Virtualization in Networks

Prof. T. Venkatesh
Dept of CSE, IIT Guwahati

What is Virtualization?

- Fundamental component of cloud computing and software defined networking
- Allows creation of isolated execution environment for multi-user environments
- Basic idea: ability of a computer program (software and hardware) to emulate an executing environment separate from the one that hosts such programs.
- Layer of indirection to run multiple software instances of a function on single hardware

Network Virtualization

- Physical components that make up a network are virtualized
- Combine hardware and software network resources, as well as network functionality into a software-based virtual network

- External network virtualization
  - Combine many networks, or parts of networks, into a virtual unit (VLANs)
- Internal network virtualization
  - Provide network switch-like functionality to the VMs on a single system (vSwitch)

Network Virtualization

- Desirable properties of network virtualization :
  - Scalability
    - Easy to extend resources in need
    - Administrator can dynamically create or delete virtual network connection
  - Resilience
    - Recover from the failures
    - Virtual network will automatically redirect packets by redundant links
  - Security
    - Increased path isolation and user segmentation
    - Virtual network should work with firewall software
  - Availability
    - Access network resource anytime

Network Virtualization

- External network virtualization in different layers :
  - Layer 2
    - Use some tags in MAC address packet to provide virtualization.
    - Example, VLAN.
  - Layer 3
    - Use some tunnel techniques to form a virtual network.
    - Example, VPN.
  - Layer 4 or higher
    - Build up some overlay network for some application.
    - Example, P2P.

Network Virtualization

- Internal network virtualization in different layers :
  - Layer 2
    - Implement virtual L2 network devices, such as switch, in hypervisor.
    - Example, Linux TAP driver + Linux bridge.
  - Layer 3
    - Implement virtual L3 network devices, such as router, in hypervisor.
    - Example, Linux TUN driver + Linux bridge + iptables.
  - Layer 4 or higher
    - Layer 4 or higher layers virtualization is usually implemented in guest OS.

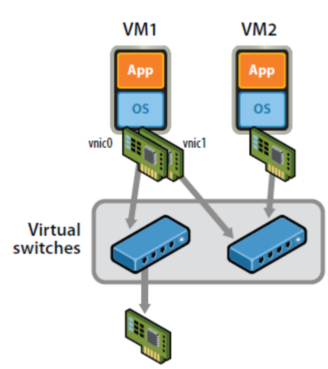
Internal Network Virtualization

- Internal network virtualization
  - A single system is configured with virtual machines, combined with hypervisor control programs or pseudo-interfaces such as the VNIC, to create a "network in a box".
  - This solution improves overall efficiency of a single system by isolating applications to separate containers and/or pseudo interfaces.
  - Virtual machine and virtual switch :
    - The VMs are connected logically to each other so that they can send data to and receive data from each other.
    - Each virtual network is serviced by a single virtual switch.
    - A virtual network can be connected to a physical network by associating one or more network adapters (uplink adapters) with the virtual switch.

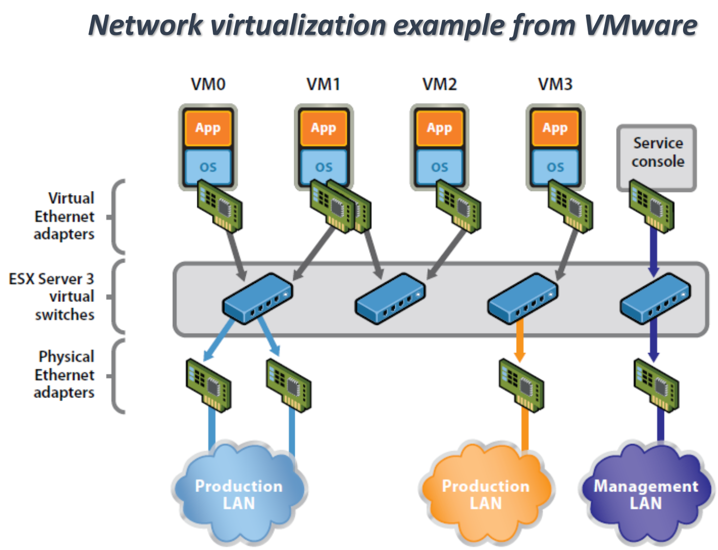
Internal Network Virtualization

- Properties of virtual switch
  - A virtual switch works much like a physical Ethernet switch.
  - It detects which VMs are logically connected to each of its virtual ports and uses that information to forward traffic to the correct virtual machines.

- Typical virtual network configuration
  - Communication network
    - Connect VMs on different hosts
  - Storage network
    - Connect VMs to remote storage system
  - Management network
    - Individual links for system administration

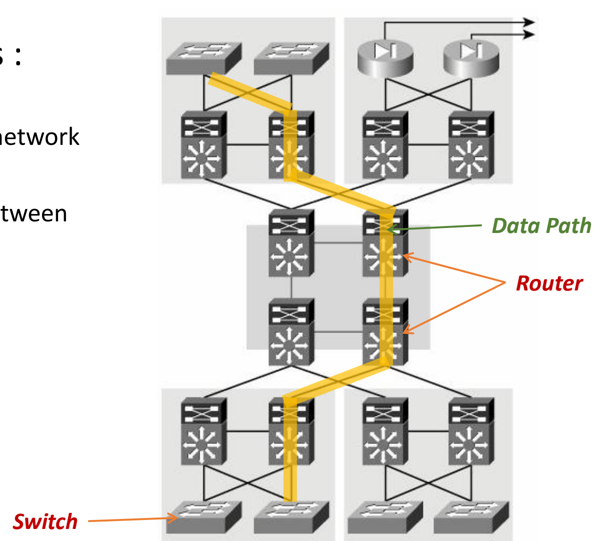


Internal Network Virtualization



External Network Virtualization

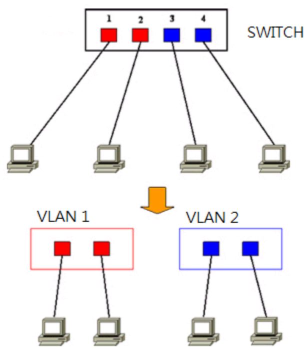
- Two virtualization components :
  - Device virtualization
    - Virtualize physical devices in the network
  - Data path virtualization
    - Virtualize communication path between network access points



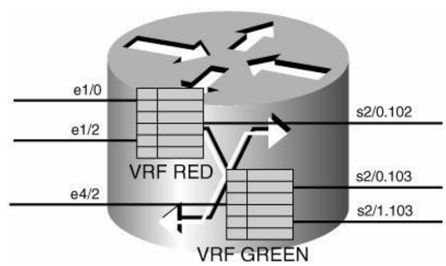
10

Network Virtualization

- Device virtualization
  - Layer 2 solution
    - Divide physical switch into multiple logical switches.



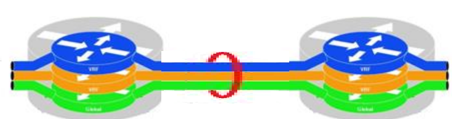
- Layer 3 solution
  - VRF technique
    (Virtual Routing and Forwarding)
  - Emulate isolated routing tables within one physical router.



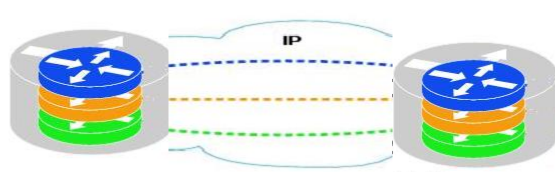
11

Network Virtualization

- Data path virtualization
  - Hop-to-hop case
    - Consider the virtualization applied on a single hop data-path.
  - Hop-to-cloud case
    - Consider the virtualization tunnels allow multi-hop data-path.



L2 based labeling allows single hop data path virtualization



Tunnels allow multi-hop data path virtualization

12

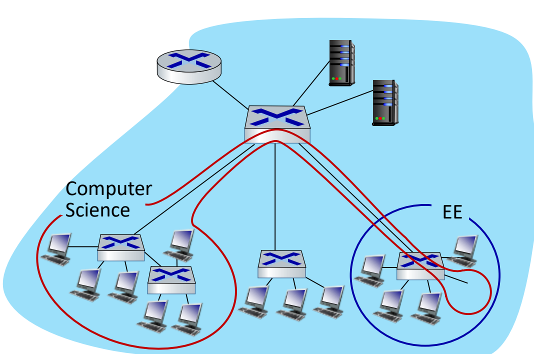
Network Virtualization

- Protocol approach
  - Protocols usually use for data-path virtualization.
  - Three implementations
    - **802.1Q** – implement hop to hop data-path virtualization
    - **MPLS (Multiprotocol Label Switch)** – implement router and switch layer virtualization
    - **GRE (Generic Routing Encapsulation)** – implement virtualization among wide variety of networks with tunneling technique.

13

Virtual LANs (VLANs): motivation

Q: what happens as LAN sizes scale, users change point of attachment?



single broadcast domain:

- *scaling:* all layer-2 broadcast traffic (ARP, DHCP, unknown MAC) must cross entire LAN
- efficiency, security, privacy, efficiency issues

administrative issues:

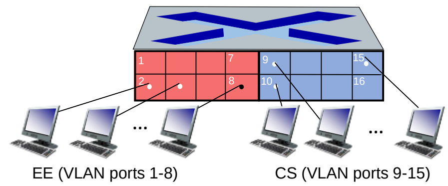
- CS user moves office to EE - *physically* attached to EE switch, but wants to remain *logically* attached to CS switch

Port-based VLANs

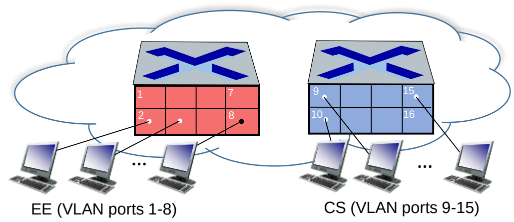
Virtual Local Area Network (VLAN)

switch(es) supporting VLAN capabilities can be configured to define multiple *virtual* LANS over single physical LAN infrastructure.

port-based VLAN: switch ports grouped (by switch management software) so that *single* physical switch ……

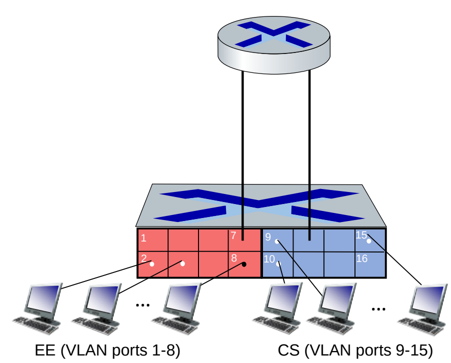


… operates as multiple virtual switches

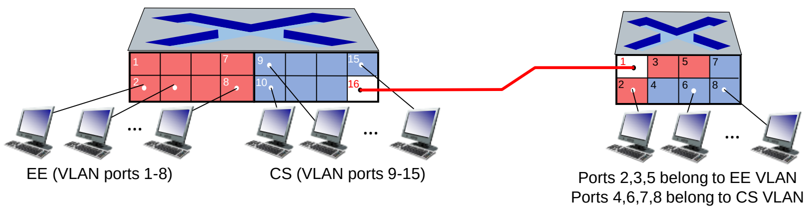


Port-based VLANs

- **traffic isolation:** frames to/from ports 1-8 can *only* reach ports 1-8
  - can also define VLAN based on MAC addresses of endpoints, rather than switch port
- **dynamic membership:** ports can be dynamically assigned among VLANs
- **forwarding between VLANS:** done via routing (just as with separate switches)
  - in practice vendors sell combined switches plus routers



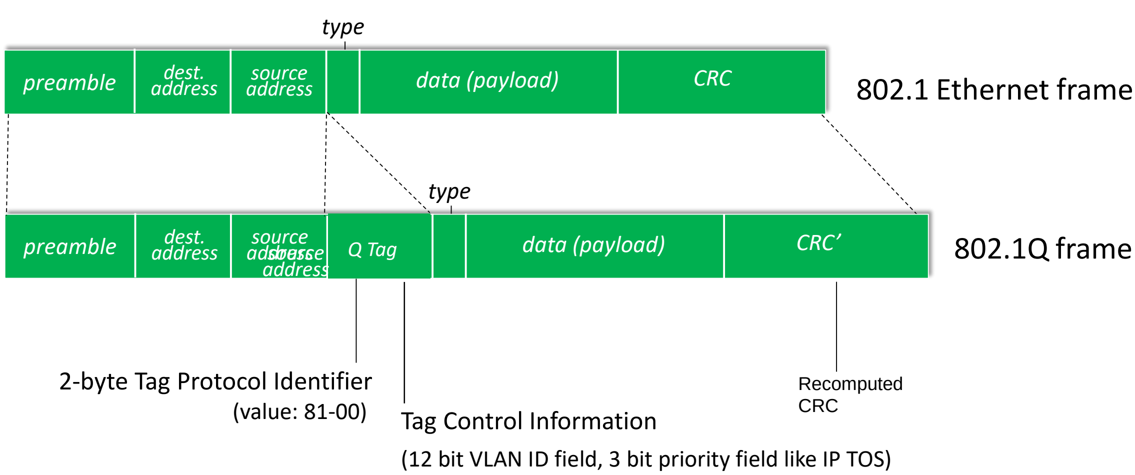
VLANs spanning multiple switches



trunk port: carries frames between VLANS defined over multiple physical switches

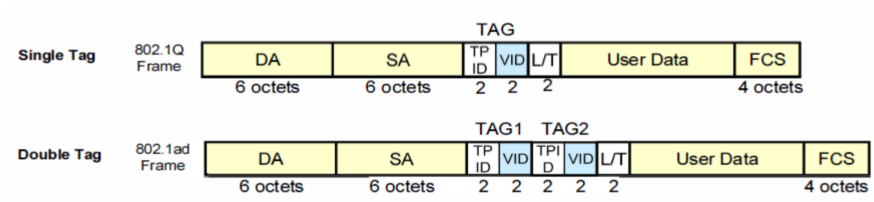
- frames forwarded within VLAN between switches can't be vanilla 802.1 frames (must carry VLAN ID info)
- 802.1q protocol adds/removed additional header fields for frames forwarded between trunk ports

802.1Q VLAN frame format



Q-in-Q Encapsulation

- Use the existing Ethernet header (802.1ad) but forward according to ingress port and VLAN id, not MAC address
- Add tags if required (label stacking)
- Provider inserts a service VLAN tag, VLAN translation changes VLANs using a table
- Forwarding decision based on single or multiple VLAN ids with link-local scope
- Replace flooding and learning bridges with switched VLAN traffic



VC Switching-in a Nutshell

"source-to-dest path behaves much like telephone circuit"

- performance-wise
- network actions along source-to-dest path

- call setup, teardown for each call *before* data can flow
- each packet carries VC identifier (not destination host address)
- *every* router on source-dest path maintains "state" for each passing connection
- link, router resources (bandwidth, buffers) may be *allocated* to VC

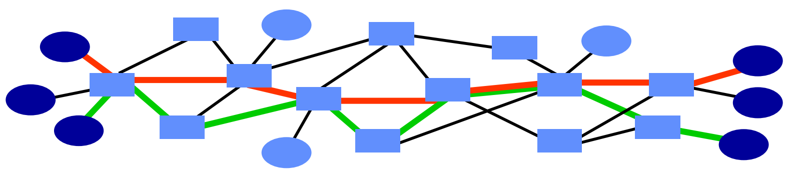
A VC consists of:

1. Path from source to destination
2. VC numbers, one number for each link along path
3. Entries in forwarding tables in routers along path

- VC numbers are configured as a part of forwarding table
  - Signaling protocol used to configure VC paths

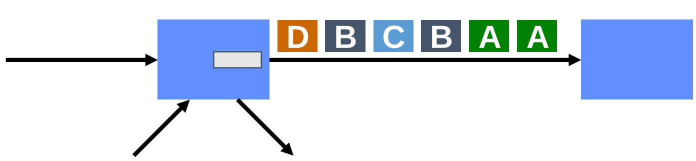
Virtual Circuits

- Each wire carries many "virtual" circuits.
  - Forwarding based on virtual circuit (VC) identifier
    - IP header: src, dst, etc.
    - Virtual circuit header: just "VC"
  - A path through the network is determined for each VC when the VC is established
  - Use statistical multiplexing for efficiency
- Can support wide range of quality of service.
  - No guarantees: best effort service
  - Weak guarantees: delay < 300 msec, …
  - Strong guarantees: e.g. equivalent of physical circuit



Similarities with packet switching

- "Store and forward" communication based on an address.
  - Address is either the destination address or a VC identifier
- Must have buffer space to temporarily store packets.
  - E.g. multiple packets for some destination arrive simultaneously
- Multiplexing on a link is similar to time sharing.
  - No reservations: multiplexing is statistical, i.e. packets are interleaved without a fixed pattern
  - Reservations: some flows are guaranteed to get a certain number of "slots"



Differences from packet switching

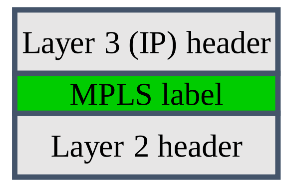
- Circuit switching:
  - Uses short connection identifiers to forward packets
  - Switches know about the connections so they can more easily implement features such as quality of service
  - Virtual circuits form basis for traffic engineering: VC identifies long-lived stream of data that can be scheduled
- Packet switching:
  - Use full destination addresses for forwarding packets
  - Can send data right away: no need to establish a connection first
  - Switches are stateless: easier to recover from failures
  - Adding QoS is hard
  - Traffic engineering is hard: too many packets!

VC setup: Permanent VCs and Switched VCs

- Permanent vs. Switched virtual circuits (PVCs, SVCs)
- Main difference is: static vs. dynamic.
- PVCs last “a long time”
  - E.g., connect two bank locations with a direct link (really expensive!) or setup a PVC that looks like a circuit
  - Administratively configured
- SVCs is temporary
  - Setup is more like a phone call
  - SVCs dynamically set up on a “per-call” basis

Multi-Protocol Label Switching (MPLS)

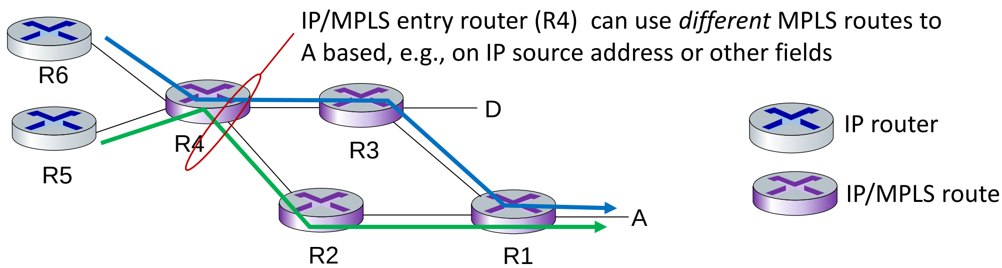
- A forwarding scheme designed to speed up IP packet forwarding (RFC 3031)
- Idea: use a fixed length label in the packet header to decide packet forwarding
- Label carried in an MPLS header between the link layer header and network layer header
  - Existing routers could act as MPLS switches just by examining the MPLS label-- no radical re-design
- MPLS tunnels used for VPNs, traffic engineering, reduced core routing table sizes
- Support any network layer protocol and link layer protocol



MPLS capable routers

- a.k.a. label-switched router
- forward packets to outgoing interface based only on label value (*don't inspect IP address*)
  - MPLS forwarding table distinct from IP forwarding tables
- *flexibility:* MPLS forwarding decisions can *differ* from those of IP
  - use destination *and* source addresses to route flows to same destination differently (traffic engineering)
  - re-route flows quickly if link fails: pre-computed backup paths

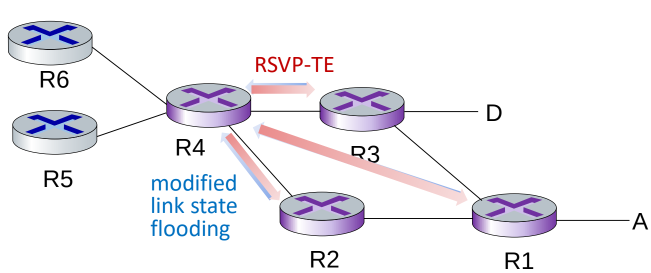
MPLS versus IP paths



- IP routing: path to destination determined by destination address alone
- MPLS routing: path to destination can be based on source *and* destination address
  - flavor of generalized forwarding (MPLS 10 years earlier)
  - *fast reroute:* precompute backup routes in case of link failure

MPLS signaling

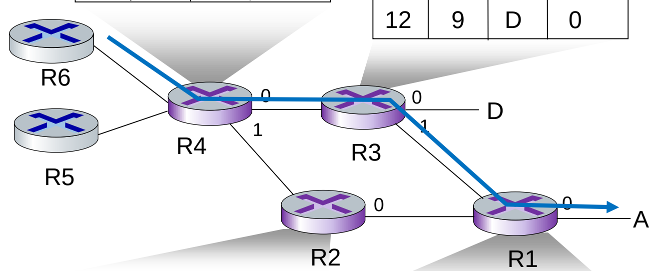
- modify OSPF, IS-IS link-state flooding protocols to carry info used by MPLS routing:
  - e.g., link bandwidth, amount of "reserved" link bandwidth
- entry MPLS router uses RSVP-TE signaling protocol to set up MPLS forwarding at downstream routers



MPLS forwarding tables

in label	out label	dest	out interface
	10	A	0
	12	D	0
	8	A	1

in label	out label	dest	out interface
10	6	A	1
12	9	D	0



in label	out label	dest	out interface
8	6	A	0

in label	out label	dest	out interface
6	-	A	0

Key Ideas

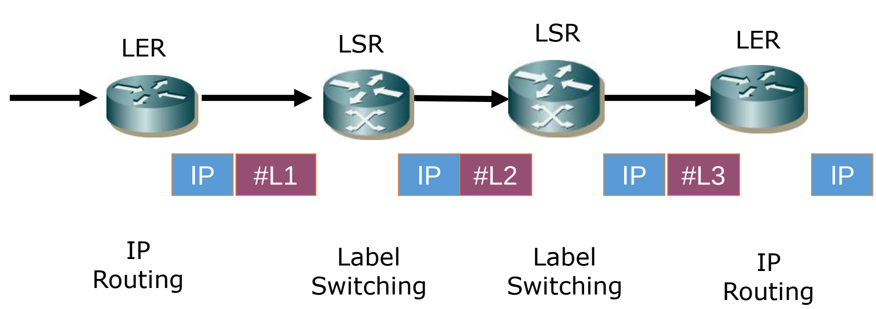
- Packets are switched, not routed, based on labels
- Labels are inserted transparently in the packet header
- Label swapping: Labels only have link-local scope
- Separation of forwarding plane and control plane
- Constraint-based routing: Traffic Engineering, Fast reroute
- Facilitate the virtual private networks (VPNs)
- Provide QoS - mapping DiffServ fields onto an MPLS label
- Establish the forwarding table
  - Link state routing protocols
    - Exchange network topology information for path selection: OSPF-TE, IS-IS-TE
  - Signaling/Label distribution protocols
    - Set up LSPs (Label Switched Path): LDP, RSVP-TE, CR-LDP

Terminology

- LSR - Routers that support MPLS are called Label Switch Router
- LER - LSR at the edge of the network is called Label Edge Router (Edge LSR)
  - Ingress LER is responsible for adding labels to unlabeled IP packets.
  - Egress LER is responsible for removing the labels.
- Label Switch Path (LSP) – the path defined by the labels through LSRs between two LERs.
- Label Forwarding Information Base (LFIB) – a forwarding table (mapping) between labels to outgoing interfaces.
- Forward Equivalent Class (FEC) – All IP packets follow the same path on the MPLS network and receive the same treatment at each node.

MPLS Operation

- At ingress LSR of an MPLS domain, an MPLS header is inserted to a packet before the packet is forwarded
  - Label in the MPLS header encodes the packet's FEC
- At subsequent LSRs
  - The label is used as an index into a forwarding table that specifies the next hop and a new label.
  - The old label is replaced with the new label, and the packet is forwarded to the next hop.
- Egress LSR strips the label and forwards the packet to final destination based on the IP packet header



Forwarding Equivalence Class

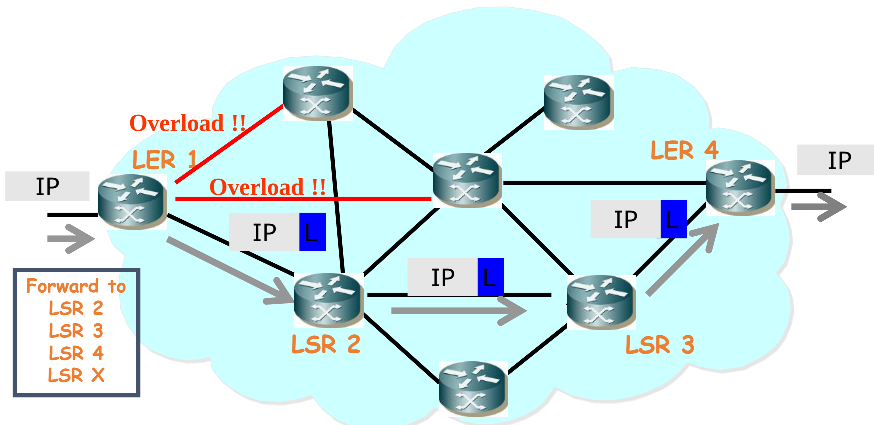
- Forwarding Equivalence Class (FEC): A subset of packets that are all treated the same way by an LSR
- A packet is assigned to an FEC at the ingress of an MPLS domain
- A packet's FEC can be determined by one or more of the following:
  - Source and/or destination IP address
  - Source and/or destination port number
  - Protocol ID
  - Differentiated services code point
  - Incoming interface
- A particular PHB (scheduling and discard policy) can be defined for a given FEC

33

MPLS Applications

- Traffic Engineering
- Virtual Private Network
- Quality of Service (QoS)
- Faster Restoration

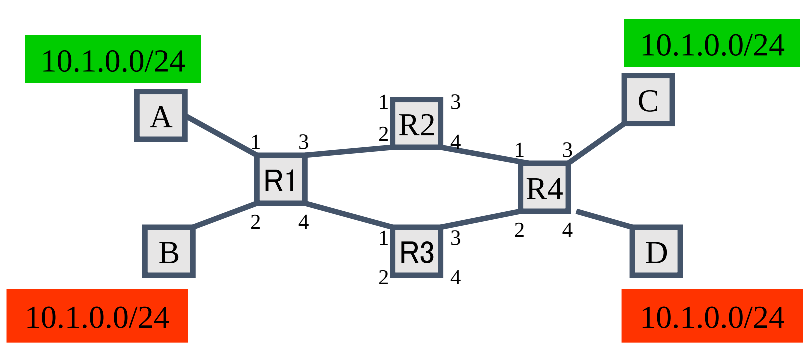
MPLS – Traffic Engineering



- **End-to-End forwarding decision determined by ingress node.**
- **Enables Traffic Engineering**

MPLS-based VPN

- One of most popular MPLS applications is the implementation of VPN.
- Using label (instead of IP address) to interconnect multiple sites over a carrier's network. Each site has its own private IP address space.
- Different VPNs may use the same IP address space.



MPLS and QoS

- An important proposed MPLS capability is quality of service (QoS) support.
- QoS mechanisms:
  - Pre-configuration based on physical interface
  - Classification of incoming packets into different classes
  - Classification based on network characteristics (such as congestion, throughput, delay, and loss)
- A label corresponding to the resultant class is applied to the packet.
- Labeled packets are handled by LSRs in their path without needing to be reclassified.
- MPLS enables simple logic to find the state that identifies how the packet should be scheduled.
- The exact use of MPLS for QoS purposes depends a great deal on how QoS is deployed.
- Support various QoS protocols, such as IntServ, DiffServ, and RSVP.