Internetworking: Global Internet and MPLS

Hui Chen, Ph.D.
Dept. of Engineering & Computer Science
Virginia State University
Petersburg, VA 23806

Acknowledgements

- Some pictures used in this presentation were obtained from the Internet
- The instructor used the following references
  - Larry L. Peterson and Bruce S. Davie, Computer Networks: A Systems Approach, 5th Edition, Elsevier, 2011
  - Andrew S. Tanenbaum, Computer Networks, 5th Edition, Prentice-Hall, 2010
  - James F. Kurose and Keith W. Ross, Computer Networking: A Top-Down Approach, 5th Ed., Addison Wesley, 2009
  - Larry L. Peterson's (http://www.cs.princeton.edu/~llp/) Computer Networks class web site

Outline

- Challenges
  - Existences of large number of “AS’s”
  - Scale of the Global Internet
- Paradigm shift
  - Evolution of the Internet
  - EGP → BGP
- EGP and BGP
- IGP, iBGP, and eBGP: Integrating Interdomain and Intradomain routing
- VPN, tunnels, and MPLS

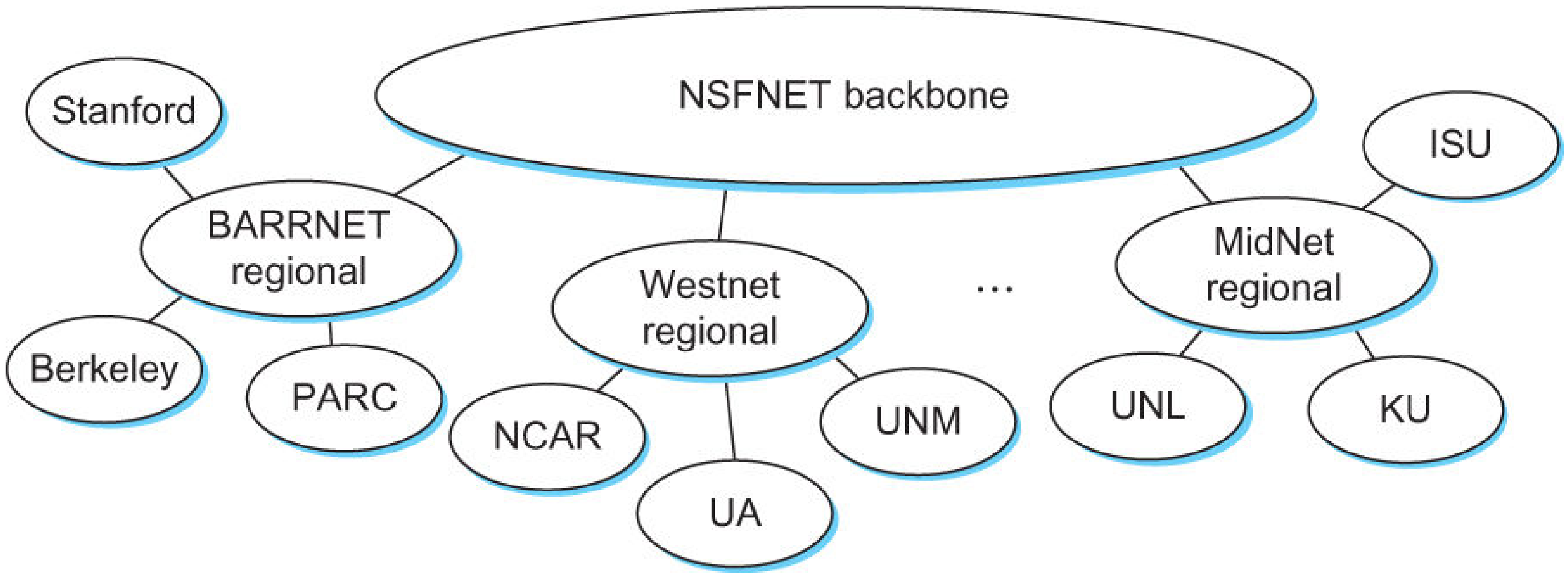
Problem

- Scale to global Internet
  - How do we build a routing system that can handle hundreds of thousands of networks and billions of end nodes?
  - How to handle address space exhaustion of IPv4?
    - IPv6 (in later lectures)
  - How to enhance the functionalities of Internet?

Evolution of the Global Internet

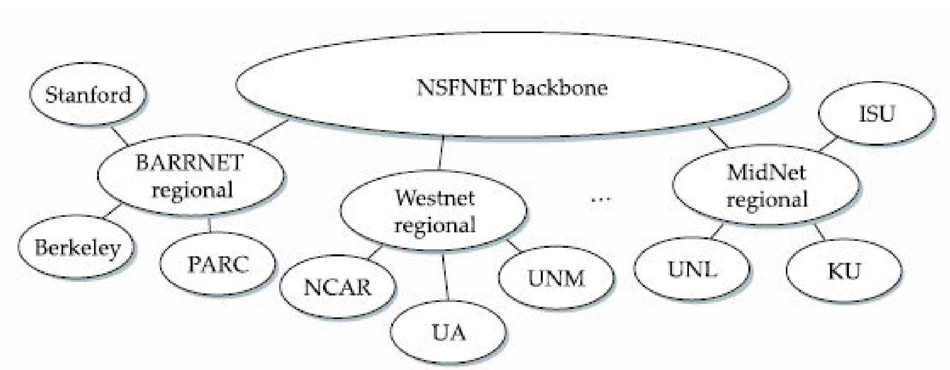
- Tree structure in 1990
- Non-tree structure today
  - Simple multi-provider Internet
  - Richly interconnected set of networks, mostly operated by private companies

Tree structure of the Internet in 1990



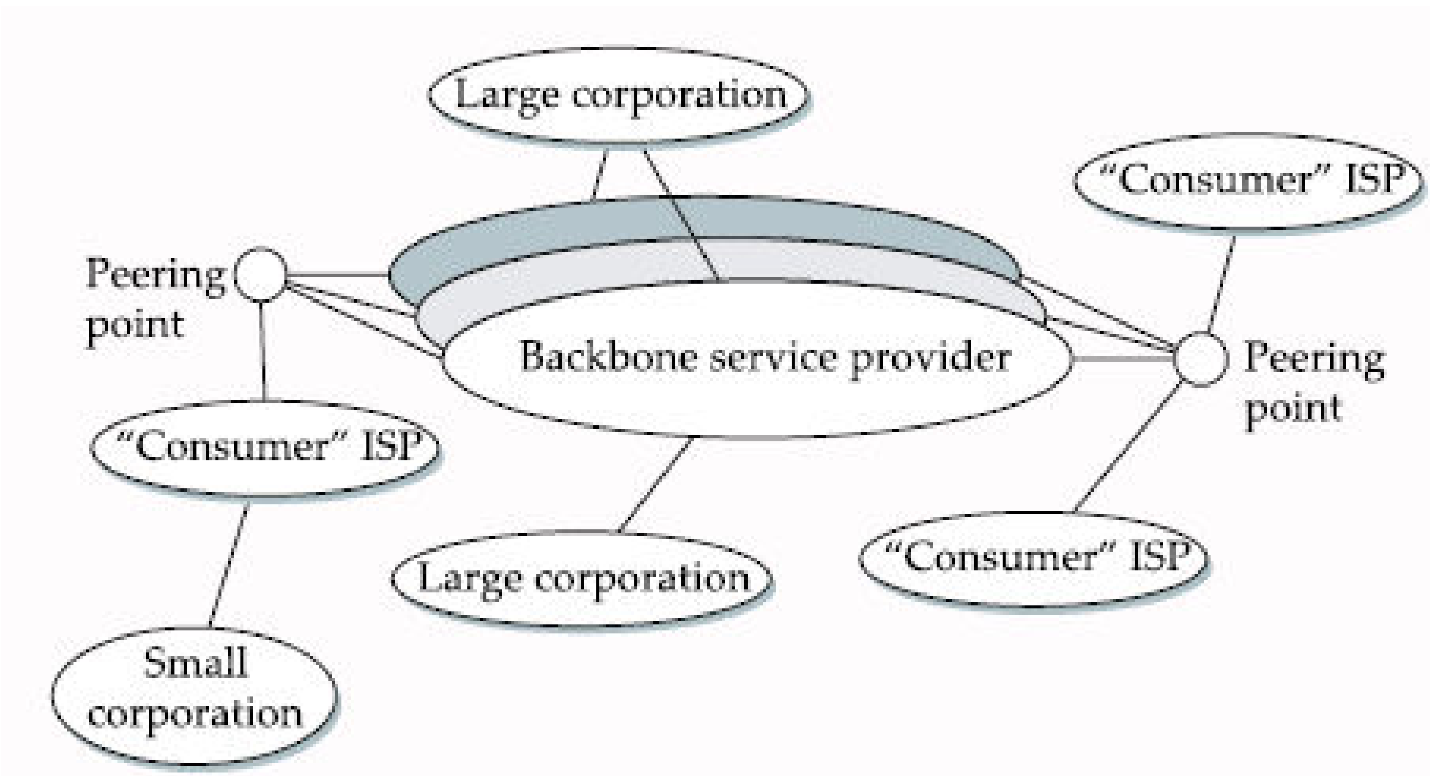
The Internet in 1990

- Hierarchical manner structures
  - Backbone network → regional networks/providers → end users
  - Many administrative independent entities: each entity decides what is the best for itself (routing algorithms, cost metrics etc)
  - Each provider is usually a single autonomous system (AS)
- Problems
  - Scalability of Routing: minimize the number of networks
  - Address utilization: every host needs an IP address



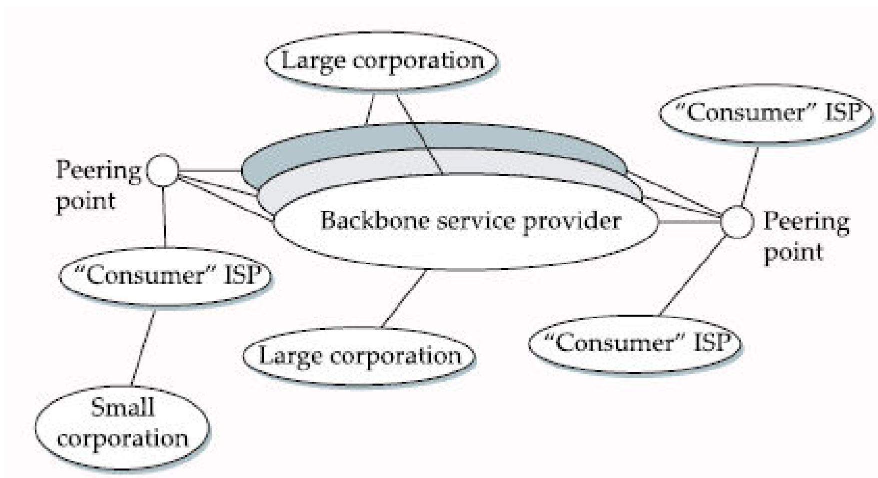
The Internet Today

- A simple multi-provider internet

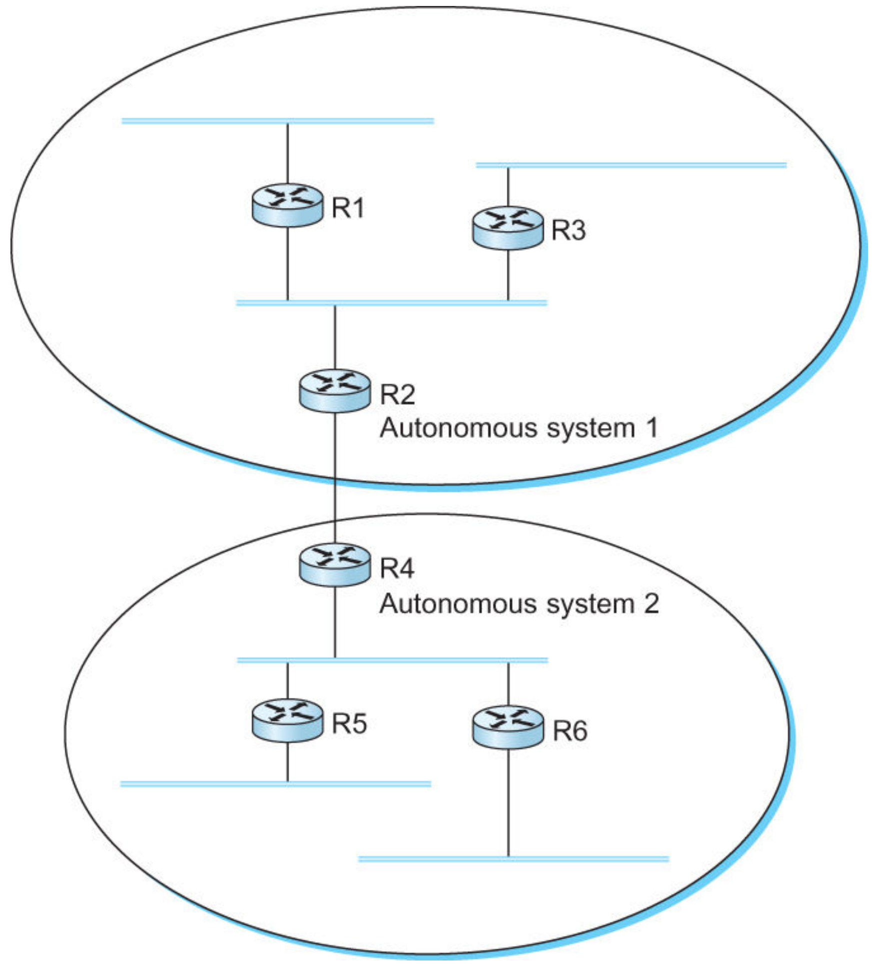


The Internet Today

- Very complex, difficult to discern much structure
  - BGP assumes that the Internet is an arbitrarily interconnected set of AS's
  - Consists of multiple backbone networks (a.k.a., service providers networks)
    - Backbone example: https://www.sprint.net/
  - Run by private companies
- Connected in arbitrary ways (the point they connect is called a peering point)



Network with Two Autonomous Systems



Scale of the Global Internet

- Using the number of AS's as a metric
  - http://as-rank.caida.org/

Inter- and Intra-Domain Routing

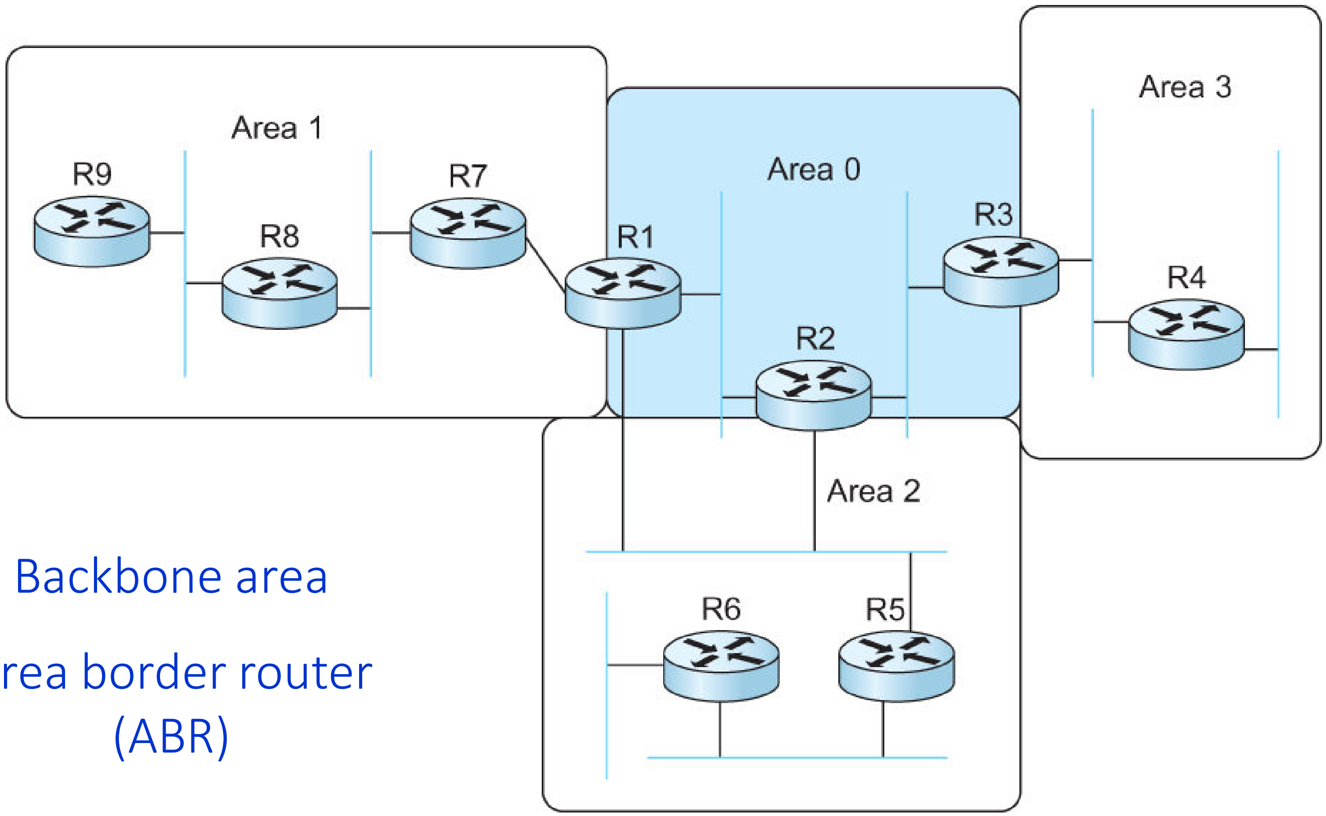
- Idea: Provide an additional way to hierarchically aggregate routing information in a large internet.
  - Improves scalability
- Divide the routing problem in two parts:
  - *Routing within a single autonomous system (intradomain routing)*
  - *Routing between autonomous systems (interdomain trouing)*
- Another name for autonomous systems in the Internet is routing domains
  - Two-level route propagation hierarchy
    - Inter-domain routing protocol (Internet-wide standard)
    - Intra-domain routing protocol (each AS selects its own)

Routing in Global Internet: Challenges

- Existence of many AS's, administratively independent entities
- Autonomous systems (a.k.a., domains or routing domains)
  - an internetwork, a network, or a subnetwork under the "jurisdiction" of a single administrative entity
- Determine their own routing policies
  - Examples:
    - Routing algorithms/protocols: RIG or OSPF?
    - Metrics/costs: by hops, bandwidth, latency, or monetary terms?
    - To which AS's should a packet be forwarded: having two providers X & Y, to which one?
    - Should I carry other AS's traffic: should I forward packet coming from X to Y, or vice versa?
    - Whom do I trust?
- An AS should implement such policies without assistance from any other AS's

Routing Areas

- A (routing) domain divided into (routing) areas



Backbone area

Area border router (ABR)

Inter-domain Routing Protocols

- Evaluation of inter-domain routing
  - EGP → BGP
- Exterior Gateway Protocol (EGP)
- Border Gateway Protocol (BGP)

EGP: Exterior Gateway Protocol

- Overview
  - Did not allow for the topology to become general
    - Tree like structure: there is a single backbone and autonomous systems are connected only as parents and children and not as peers
  - Concerned with reachability, not optimal routes
- Protocol messages
  - neighbor acquisition
    - one router requests that another be its peer
    - peers exchange reachability information
  - neighbor reachability
    - one router periodically tests if the another is still reachable; exchange HELLO/ACK messages
    - uses a k-out-of-n rule: at least k of the last n messages must fail for the router to declare its neighbor down
  - routing updates
    - peers periodically exchange their routing tables (distance-vector)

Limitations of EGP

- Resembles distance vector routing
  - Updates carry lists of destinations and distances
  - Distances are NOT reliable → measures reachability
- EGP was designed to support tree topologies, not meshes
  - False routes injected by accident can have really bad consequences (black holes)
    - Example: a router advertise that other networks can be reached in 0 distances
  - Loops can easily occur
    - all is forwarding routing tables
- EGP was not designed to easily support fragmented IP packets
  - all data is assumed to fit in MTU.
- Solutions to these and other EGP problems were all manual

BGP: Border Gateway Protocol

- Assumes that the Internet is an arbitrarily interconnected set of ASs.
- Today's Internet consists of an interconnection of multiple backbone networks
  - Usually called service provider networks and operated by private companies rather than the government
- Sites are connected to each other in arbitrary ways
  - Some large corporations connect directly to one or more of the backbone, while others connect to smaller, non-backbone service providers.
  - Many service providers exist mainly to provide service to "consumers" (individuals with PCs in their homes), and these providers must connect to the backbone providers
  - Often many providers arrange to interconnect with each other at a single "peering point"
- BGP-1 developed in 1989 to address problems with EGP.

BGP-4: Border Gateway Protocol Version 4

- Assumes the Internet is an arbitrarily interconnected set of AS's.
- Local and transit traffic
- Three types of AS’s

AS Traffic Types

- Local traffic
  - starts or ends within an AS
- Transit traffic
  - passes through an AS

AS Types

- Stub AS: has a single connection to one other AS
  - carries local traffic only
- Multihomed AS: has connections to more than one AS
  - refuses to carry transit traffic
- Transit AS: has connections to more than one AS
  - carries both transit and local traffic
- Subscribers: stub AS's and multihomed AS's

AS Number

- Assigned by IANA (http://www.iana.org/)
- 16 bit integers (http://www.iana.org/go/rfc1930): was big enough
  - Only non-stub AS's need unique AS numbers
  - Non-stub AS's are generally service providers: rare
- 32 bit AS numbers are on the way (http://www.iana.org/go/rfc4893)

Goal of BGP

- The goal of Inter-domain routing is to find any path to the intended destination that is loop free
  - Concerned with *reachability* than *optimality*
  - Finding path anywhere close to optimal is considered to be a great achievement
- Why?

Goal of BGP: Why?

- Scalability: An Internet backbone router must be able to forward any packet destined anywhere in the Internet
  - Having a routing table that will provide a match for any valid IP address
- Autonomous nature of the domains
  - It is impossible to calculate meaningful path costs for a path that crosses multiple ASs
  - A cost of 1000 across one provider might imply a great path but it might mean an unacceptable bad one from another provid
- Issues of trust
  - Provider A might be unwilling to believe certain advertisements from provider B

AS in BGP

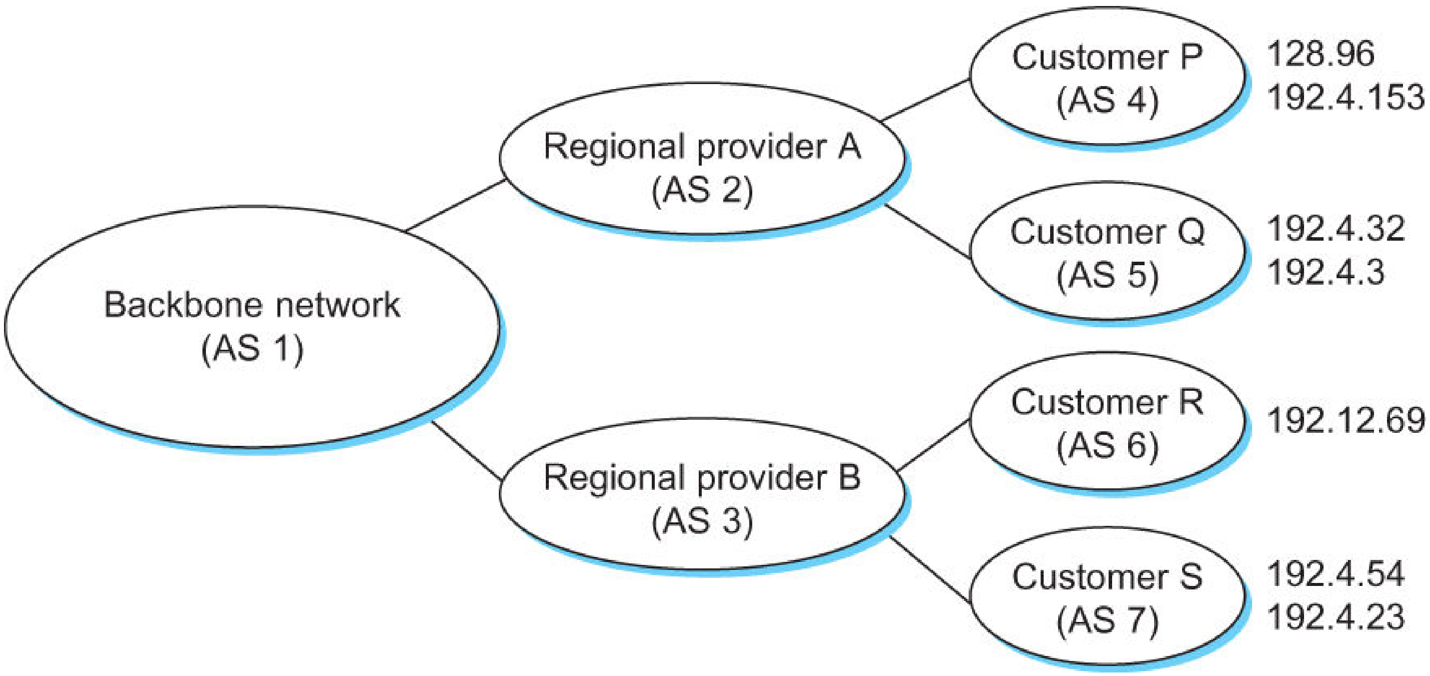
- Each AS has:
  - One or more border routers
    - handles inter-AS traffic
  - One BGP *speaker* that advertises:
    - local networks
    - other reachable networks (transit AS only)
    - gives *path* information
  - In addition to the BGP speakers, the AS has one or more border "gateways" which need not be the same as the speakers
  - The border gateways are the routers through which packets enter and leave the AS

Routing in BGP

- Classes addresses are used since BGP-4: networks are advertised as prefix/length
- BGP goal: find loop free paths between ASs
  - It's neither a distance-vector nor a link-state protocol: entire path is advertised
  - How: since *path* information is sent
    - Example: AS 2 abandons advertisements such as <A3, A2, A4> sine use it would cause a loop
- Hard problem
  - Internet's size (~12K active ASs) means large tables in BGP routers
  - Autonomous domains mean different path metrics →Optimality is secondary goal
  - Need for flexibility

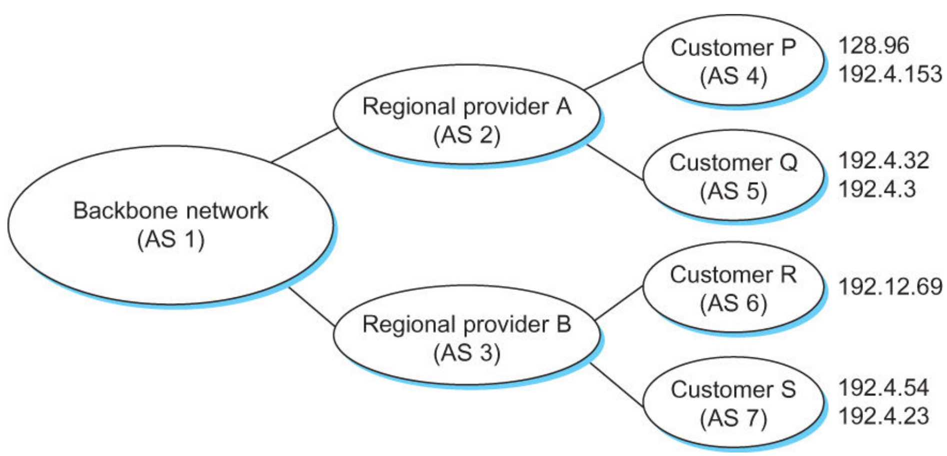
BGP: An Example

- An example network that is running BGP



BGP: An Example

- Speaker for AS2 advertises reachability to P and Q
  - Networks 128.96/16, 192.4.153/24, 192.4.32/24, and 192.4.3/24, can be reached directly from <AS2>
- Speaker for backbone (AS 1) advertises upon receiving the advertisements of the speaker of AS 2
  - Networks 128.96/16, 192.4.153/24, 192.4.32/24, and 192.4.3/24 can be reached along the path <AS1, AS2>.
- Speaker of AS 2 does not advertise anything upon receiving the above advertisement from AS 3 since the advertisement contains itself AS2 → no loop
- Speaker can cancel previously advertised paths

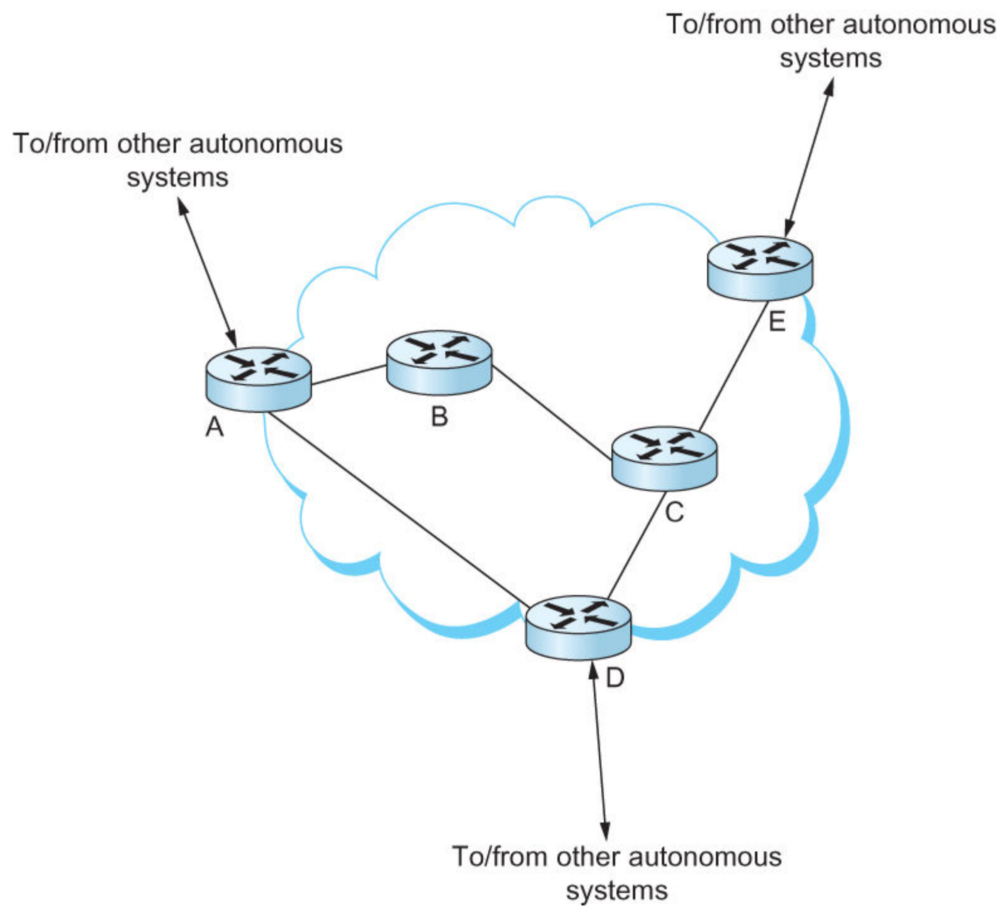


iBGP and eBGP

- Need to integrate interdomain routing and intradomain routing
  - Exterior BGP (eBGP)
    - A variant of BGP that runs between AS's
  - Interior BGP (iBGP)
    - A variant of BGP that runs on a backbone network
    - Enables any router in the AS to learn the best border router to use when sending a packet to any address
  - Intradomain domain routing protocol (IGP)
    - e.g., distance vector or link state
    - Each router that runs an IGP keeps track of how to get to each border router (within an AS)

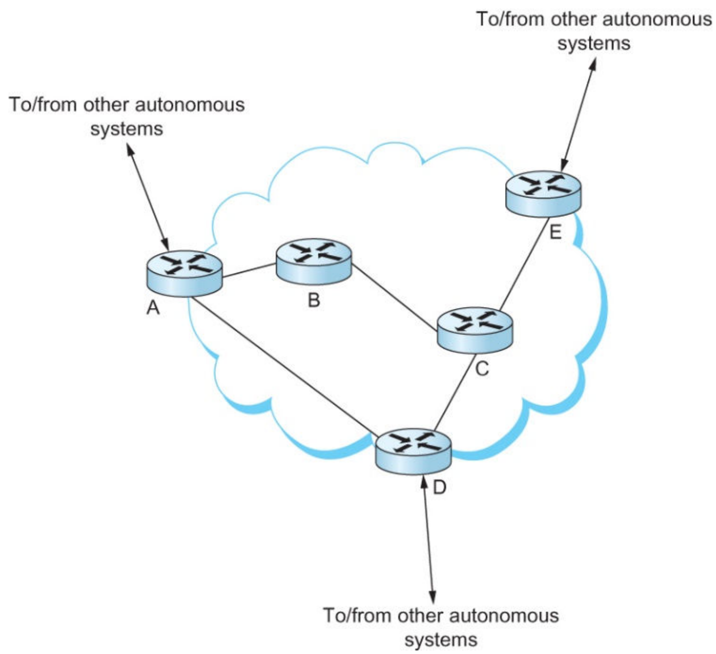
Integrating Interdomain and Intradomain Routing

- Example
  - All routers run iBGP and an intradomain routing protocol
  - Border routers (A, D, E) also run eBGP to other ASs



Routing Tables

- BGP routing table for the AS
- IGP routing table at router B
- Combined table at router B



Prefix	BGP Next Hop
18.0/16	E
12.5.5/24	A
128.34/16	D
128.69./16	A

BGP table for the AS

Router	IGP Path
A	A
C	C
D	C
E	C

IGP table for router B

Prefix	IGP Path
18.0/16	C
12.5.5/24	A
128.34/16	C
128.69./16	A

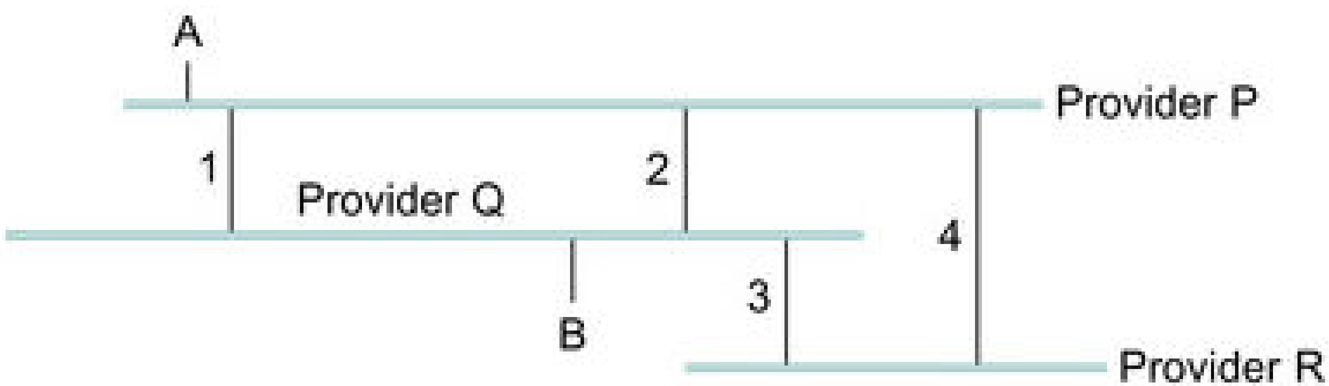
Combined table for router B

BGP-4: Some Details

- Path vectors are most important innovation in BGP
  - Enables loop prevention in complex topologies
  - If AS sees itself in the path, it will not use that path
- Routes can be aggregated
  - Based on CIDR (classless) addressing
  - Tables smaller
- Routes can be filtered
  - An AS may send a full-table view of its routing table to another AS which may only be interested in a subset.
  - Example: filter-out those not interested → tables smaller
- Runs over TCP
  - One advertisement sent will not be sent again
  - As long as no change, send "keep-alive" message → shorter than path vectors
- BGP session have only recently been made secure

Exercise L14-1

- Consider the network shown below, in which horizontal lines represent transit providers and numbered vertical lines are inter-provider links.
  - (a) How many routes to P could provider Q's BGP speakers receive?
  - (b) Suppose Q and P adopt the policy that outbound traffic is routed to the closest link to the destination's provider, thus minimizing their own cost. What paths will traffic from host A to host B and from host B to host A take?
  - (c) What could Q do to have the B → A traffic use the closer link 1?
  - (d) What could Q do to have the B → A traffic pass through R?



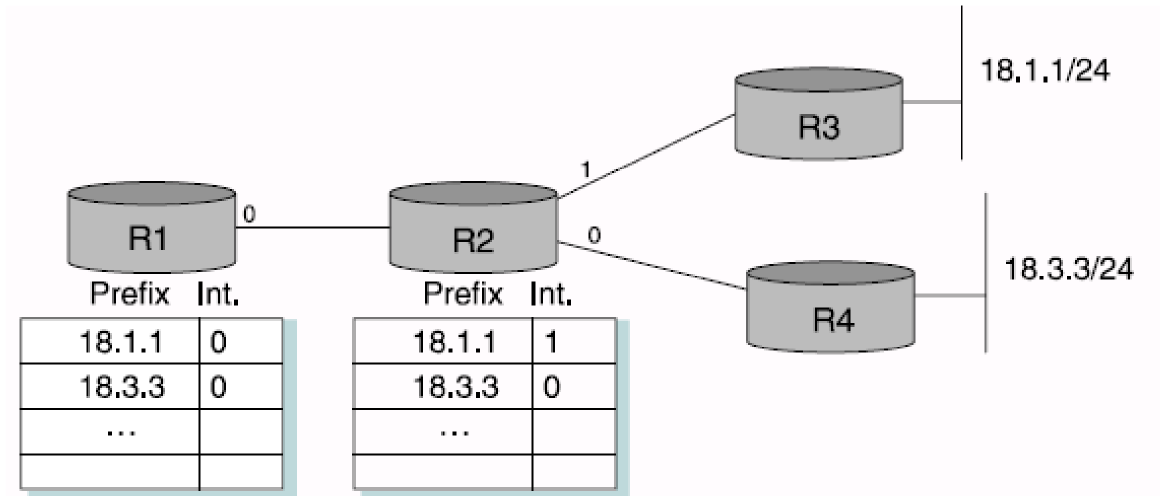
Multiprotocol Label Switching

- What is it?
- How does it work?
- Applications and benefits
- VPN and tunnels in MPLS

Multiprotocol Label Switching

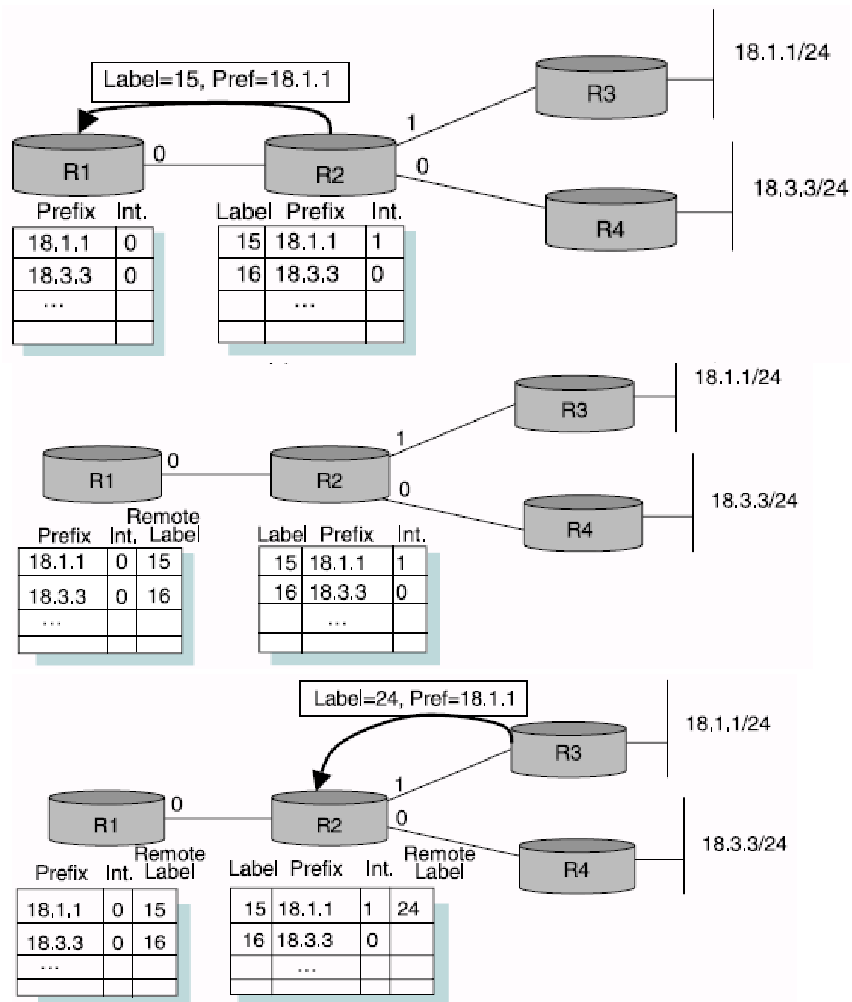
- Can be treated as a hybrid between *virtual circuits* and *datagram forwarding*
- Three main usages
  - Enable IP capabilities on non-IP devices
  - Source routing
  - Virtual private network (VPN) services

Destination-Based Forwarding in MPLS: Review of CIDR



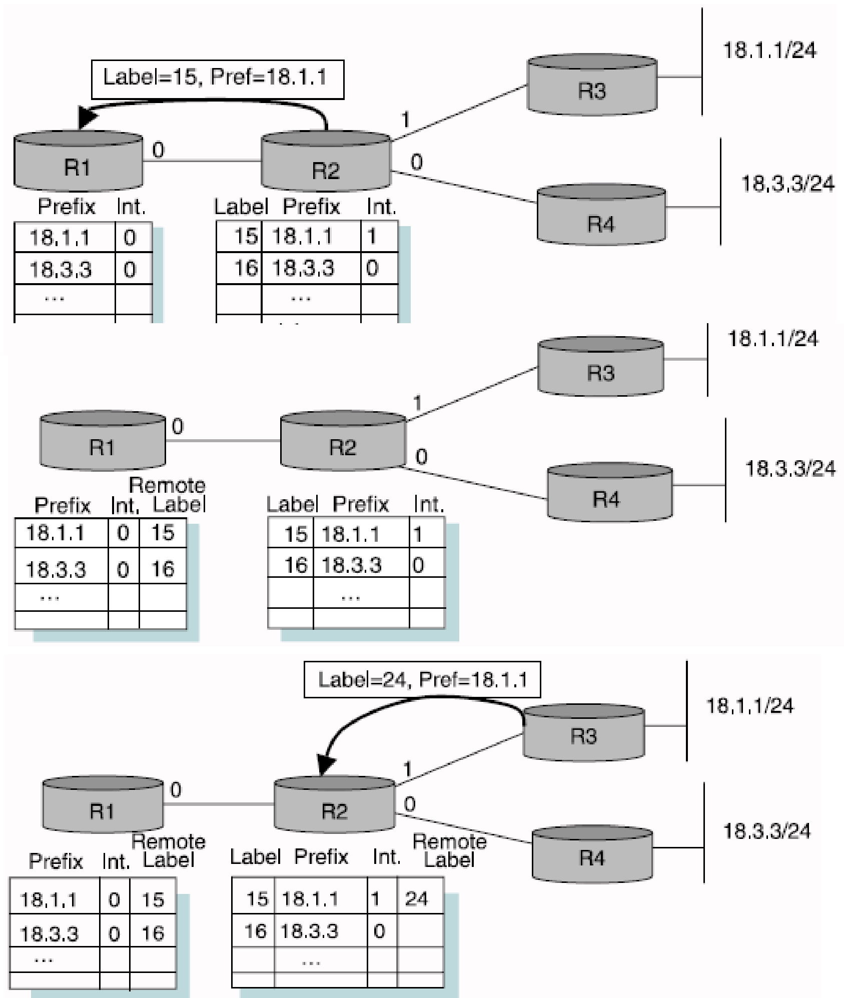
- Q: what happens when a packet destined to IP address 18.1.1.5 arrives at router R1?
  - Search the table for the longest matching prefix at R1
  - Forward the packet to router R2
  - Search the table for the longest matching prefix at R2
  - Forward the packet to router R3
  - R3 deliver it to 18.1.1/24 and the packet arrives at the host
- Happens for each packet arrives at R1

Destination-Based Forwarding in MPLS: Label Distribution



- R2 labels rows in its routing table with labels of fixed length
- R2 sends the label-and-prefix/length pair to R1
- R1 associate label to corresponding row
- Similar to R3-to-R2 label distribution

Destination-Based Forwarding in MPLS: Label "Switching"

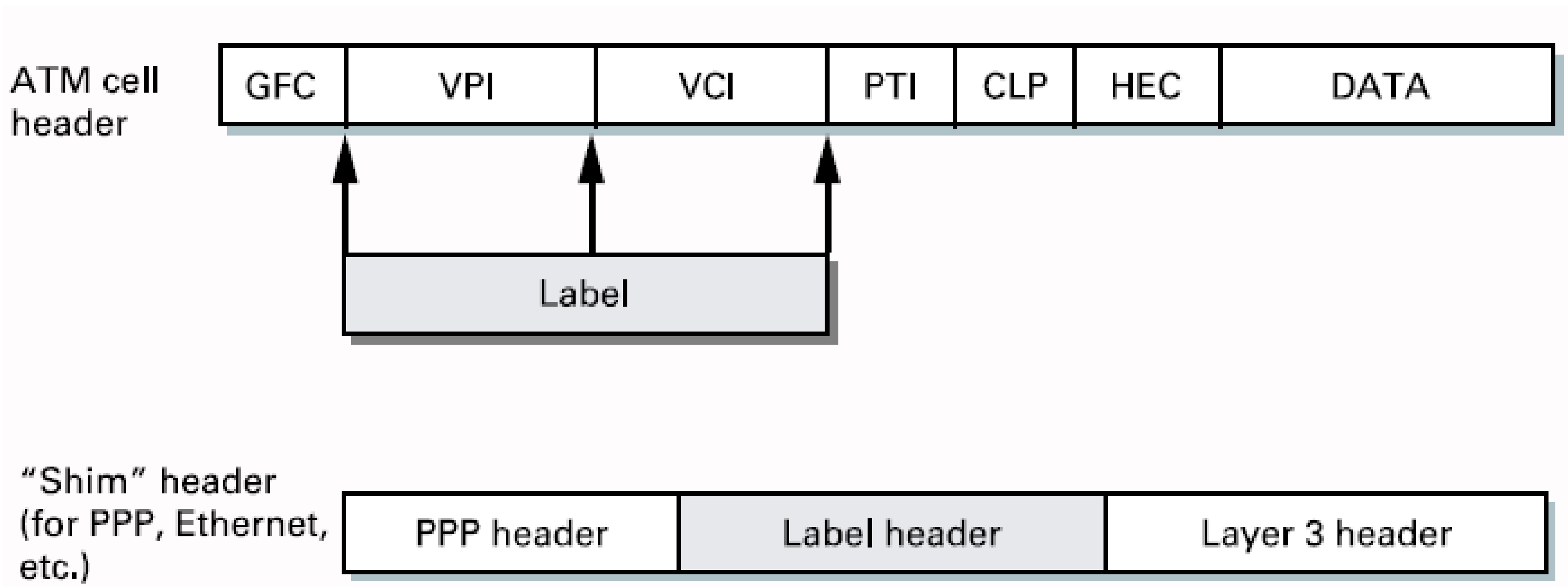


- Q: what happens when a packet destined to IP address 18.1.1.5 arrives at router R1?
- R1 is referred to as an a label edge router (LER)
- LER performs a complete IP lookup, find label 15
- Attach label 15 to the packet and sends to R2
- R2 sends to R3 based on table-lookup on label 15

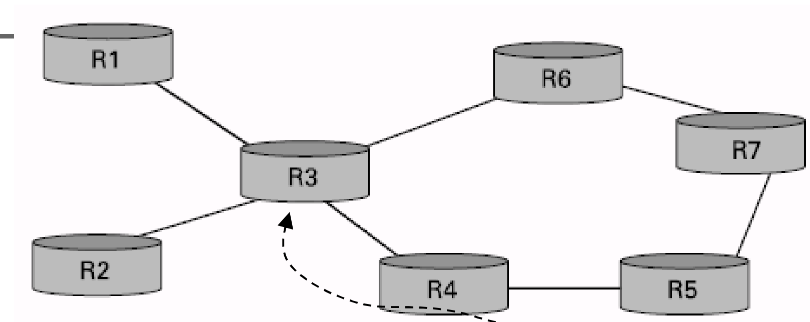
Destination-Based Forwarding in MPLS: Benefits of Label "Switching"

- Efficient table-lookup
  - Prefix/length table-lookup is expensive since we look for the longest prefix
  - Table-lookup on labels (fixed length) is very efficient (e.g., binary search)
- Labels → forwarding equivalence class (FEC)
  - A set of packets have the same treatment in terms of forwarding regardless what their IP addresses are
  - FEC can be formed using almost any criteria (not necessarily based on routing tables): all "voice" traffic can be treated as a FEC
- Enable non-IP devices to forward IP packets
  - Example: ATM supports label-swapping forwarding algorithms
    - Turn ATM into label switching routers (LSRs)
  - Can be extended to many optical switches

Destination-Based Forwarding in MPLS: How labels are attached?



Source Routing in MPLS



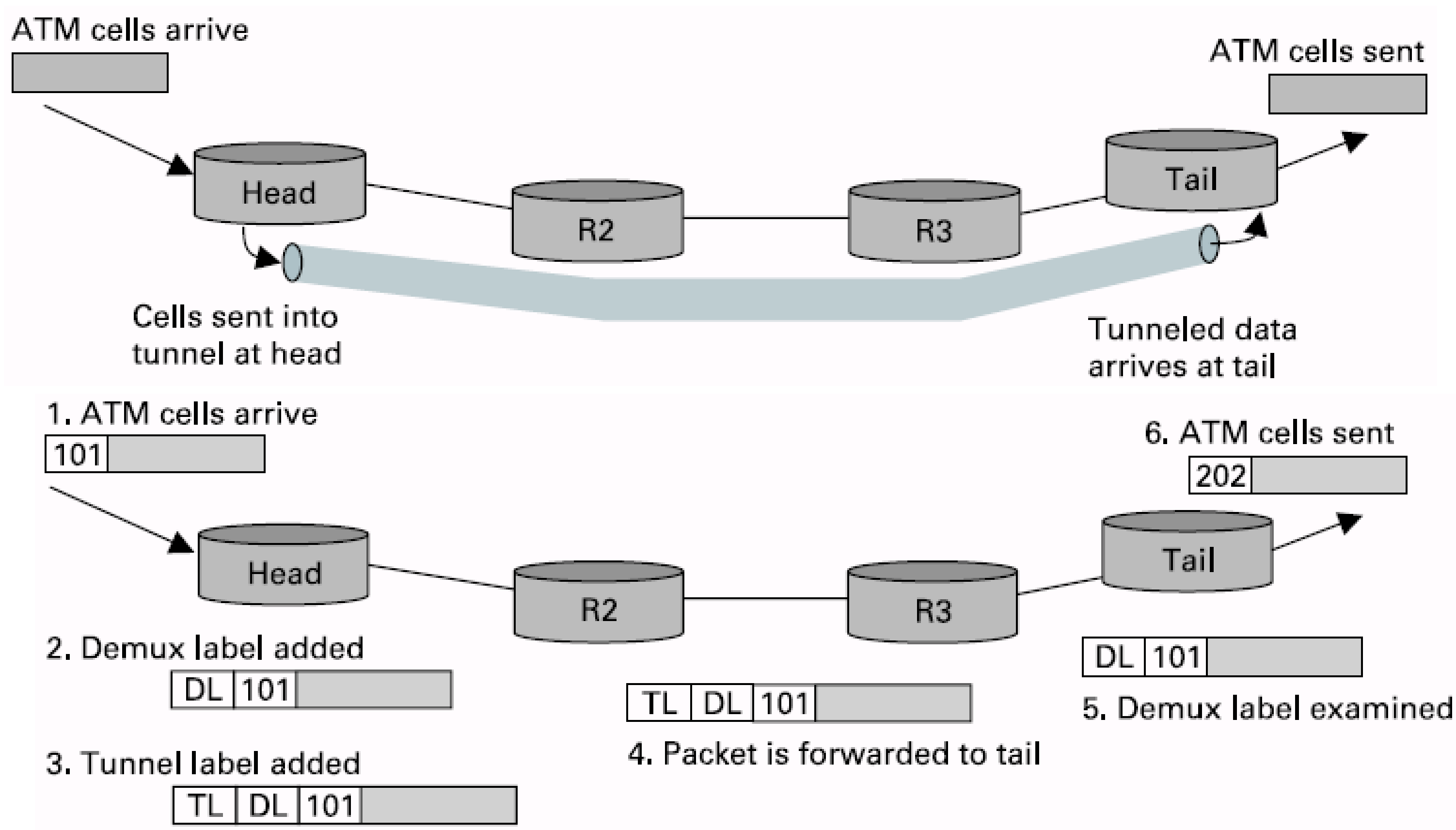
- A.k.a. explicit routing
- Example: as shown →
- In datagram forwarding
  - Forwarding based on destination address and forwarding table
  - At router R1, packets destined to R7 result in the same route
- Two FECs based on source addresses
  - FEC R1: packets forwarded by R1 to R7
    - Follow path R1-R3-R6-R7
  - FEC R2: packets forwarded by R2 to R7
    - Follow path R2-R3-R6-R7
  - Balanced load

Destination	Next Hop
R5	R4
R7	R6
…	…

Forwarding table at R3

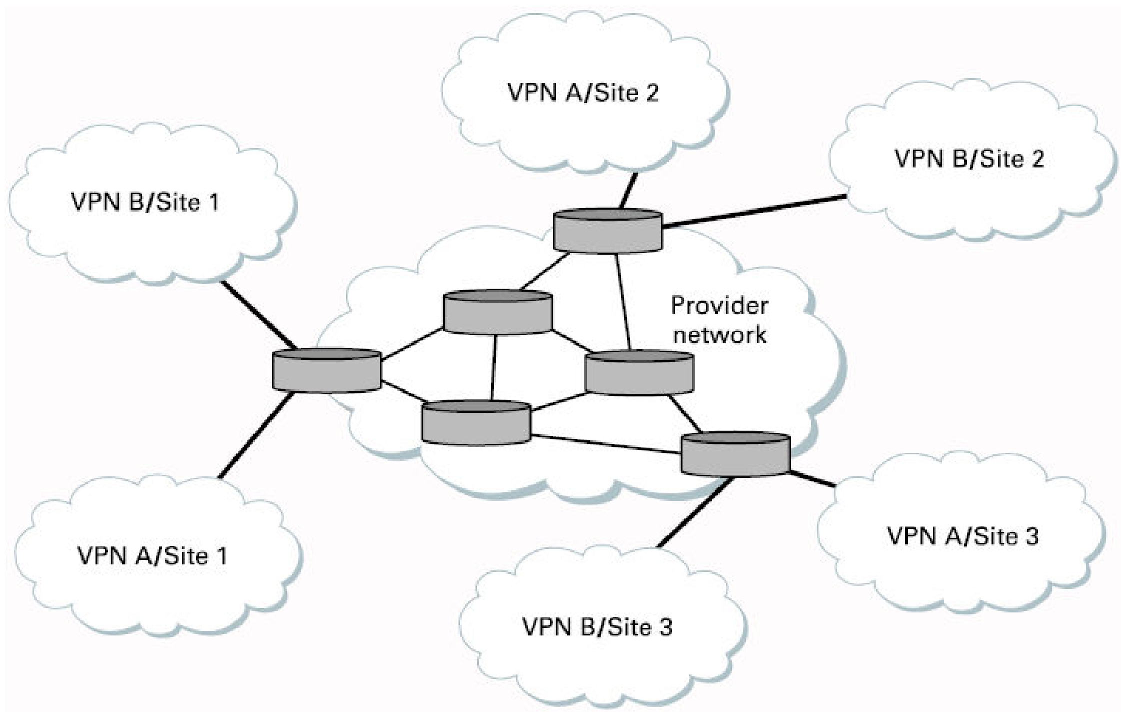
Layer 2 VPN via MPLS

- Example: emulate an ATM circuit by an MPLS tunnel



Layer 3 VPN via MPLS

- Each VPN is treated as a FEC



RFC 4364: BGP/MPLS IP Virtual Private Networks (VPNs) http://tools.ietf.org/html/rfc4364

Summary

- Challenges
  - Existences of large number of “AS’s”
  - Scale of the Global Internet
- Paradigm shift
  - Evolution of the Internet
  - EGP → BGP
- EGP and BGP
- IGP, iBGP and eBGP: Integrating Interdomain and Intradomain routing
- VPN, tunnels, and MPLS