CSC7970 – NEXT-GENERATION NETWORKING

ROUTING, INTERDOMAIN ROUTING, BGP

Instructor: Susmit Shannigrahi
sshannigrahi@tntech.edu

Tennessee TECH

Logistics

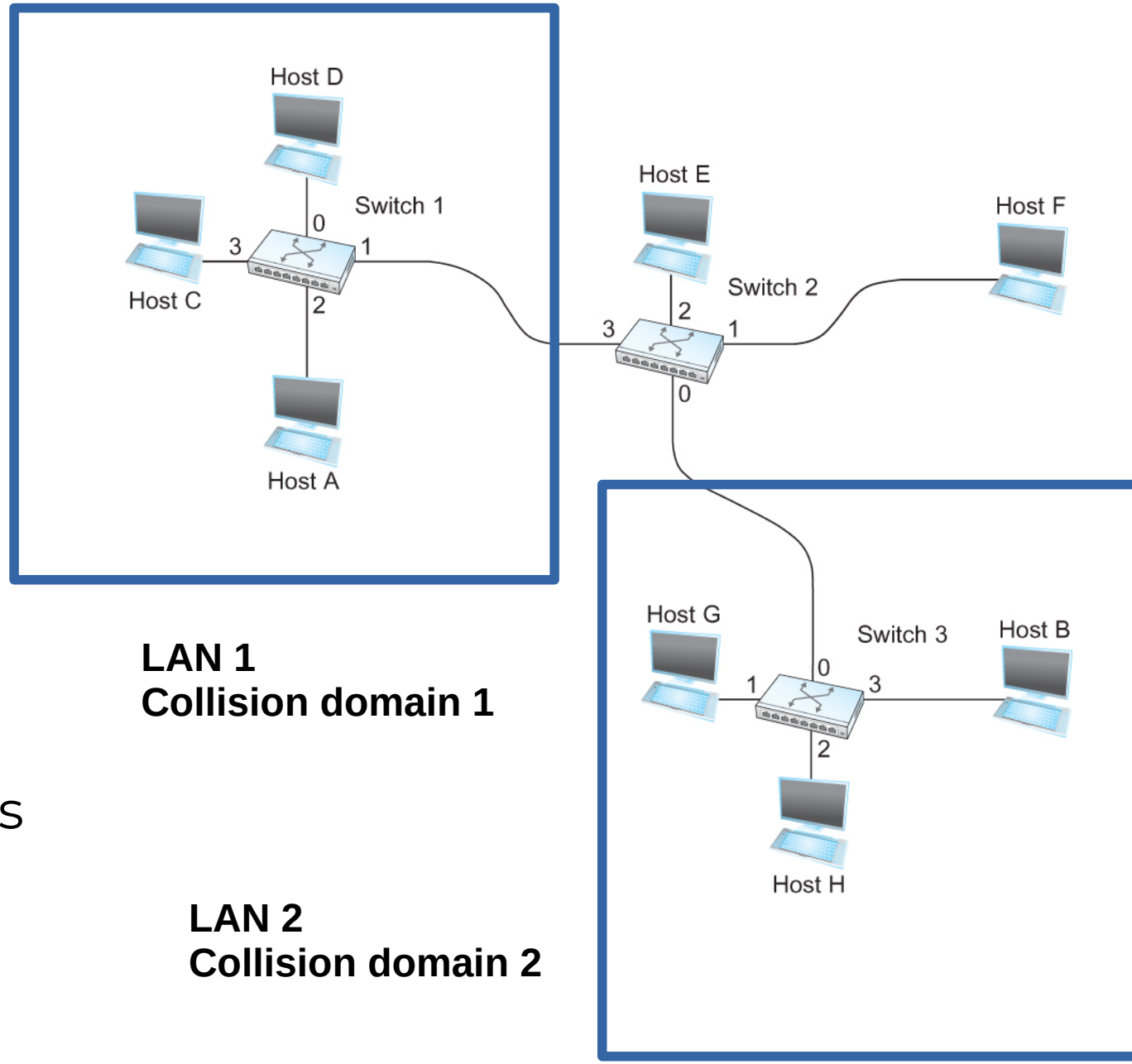
- Class website: https://tntech-ngin.github.io/csc7970
- Slack – csc7970.slack.com
- 1 Page project proposal due next week

Switching vs Routing vs Interdomain Routing

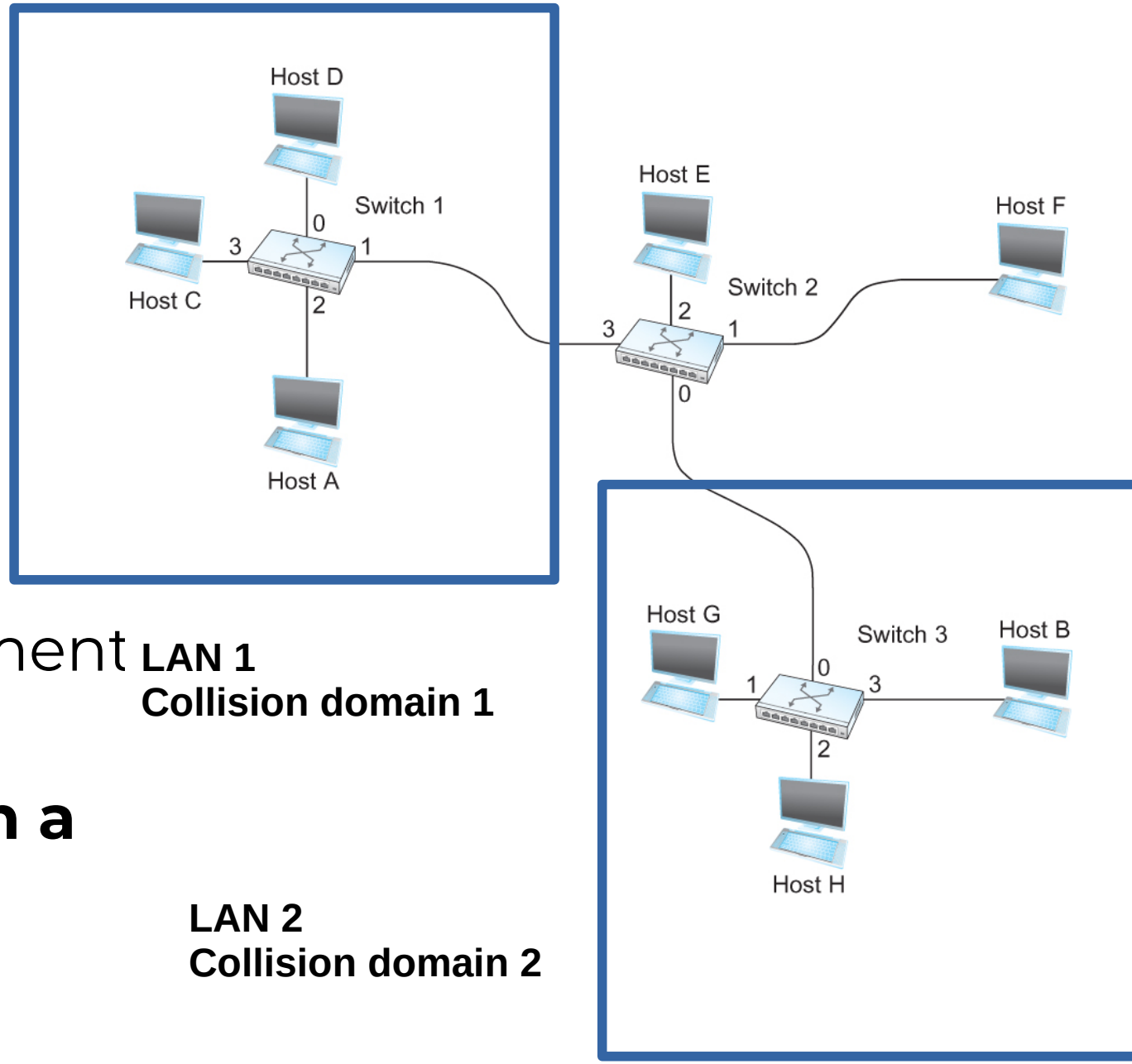
- What do you think?

Switching

- Switch
  - A mechanism to interconnect links to form a large network
  - Forward **frames**
  - Separate the collision domains
  - Filter packets between LANs
  - Connects two or more LAN segments - **Bridging**



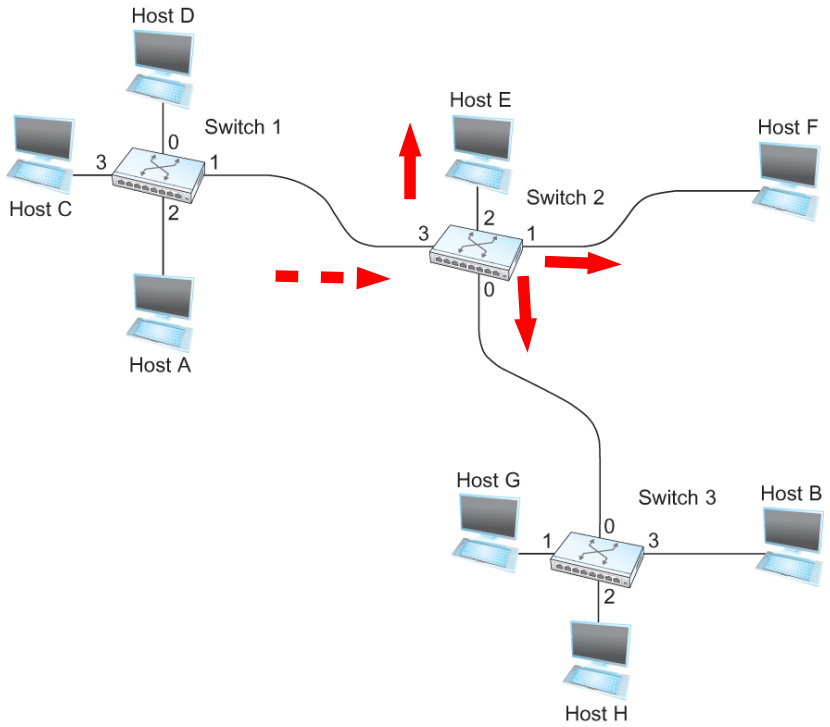
Switches are Self learning!



- No configuration needed
- Send frames to needed segment
- **How do they construct such a table?**

Switching Table

- Unknown destination → send out on all Interfaces **(flooding)**
  - **Skip the incoming interface**

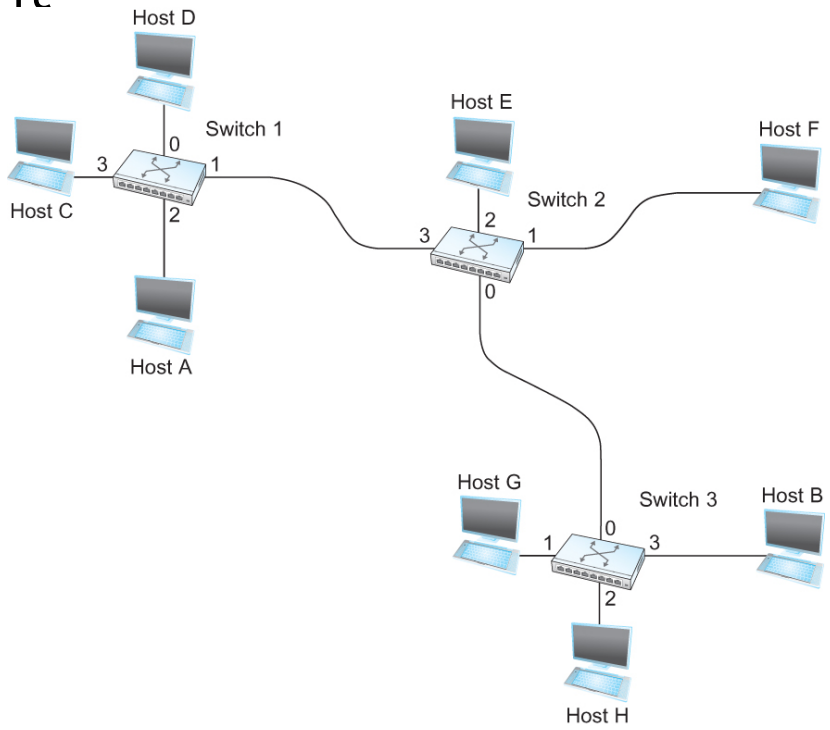


Destination, Port	
A	3
B	0
C	3
D	3
E	2
F	1
G	0
H	0

Forwarding Table for Switch 2

Switching Table Algorithm

- Create the table first!
  - **For each packet**
    - If destination address in arriving segment
      - Drop
    - If destination is in another segment
      - Forward
    - If destination unknown
      - Flood!



Routing

- How is it different than switching?
  - Switching is in the same network! Routing is between networks.
  - Switching → L2, Routing → L3

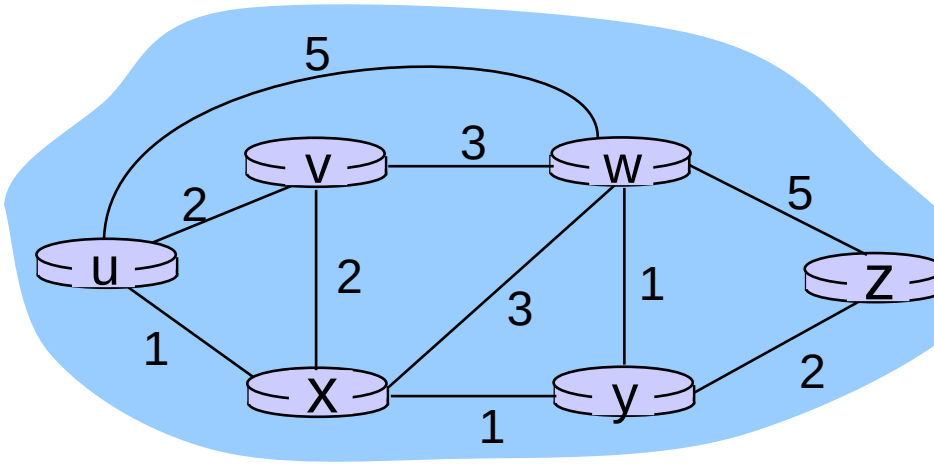
Forwarding vs Routing

- Forwarding:
  - to select an output port based on destination address and routing table
  - **Local path**

SubnetNumber	SubnetMask	NextHop
128.96.34.0	255.255.255.128	Interface 0
128.96.34.128	255.255.255.128	Interface 1
128.96.33.0	255.255.255.0	R2

- Routing:
  - process by which routing table is built
  - **End-to-end path**

Why bother?



- Quality of path affects performance
  - Longer path = more delay
- Balance path usage, avoid congested paths
- Deal with failures

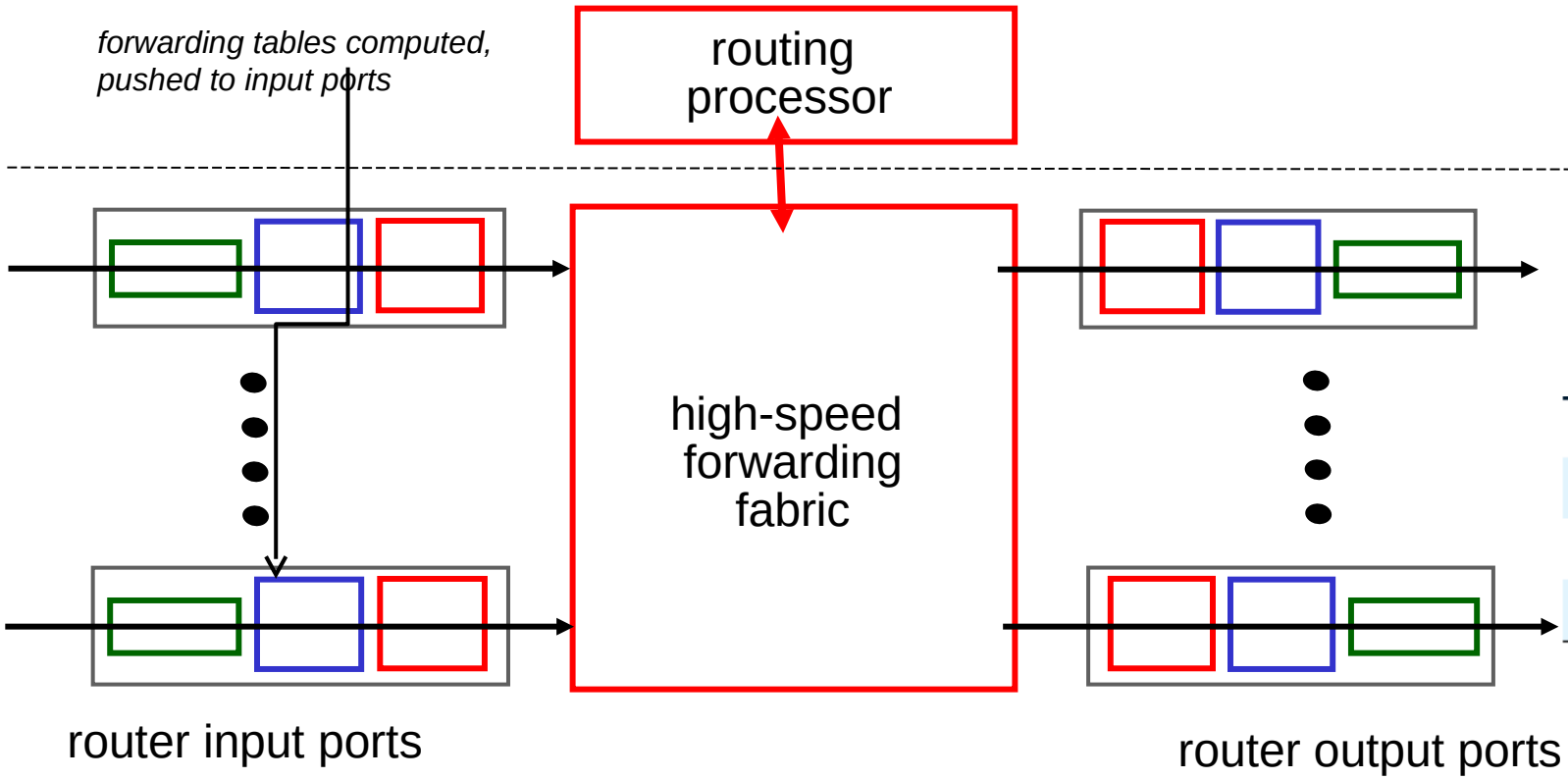
SubnetNumber	SubnetMask	NextHop
128.96.34.0	255.255.255.128	Interface 0
128.96.34.128	255.255.255.128	Interface 1
128.96.33.0	255.255.255.0	R2

Router architecture overview

Two key router functions:

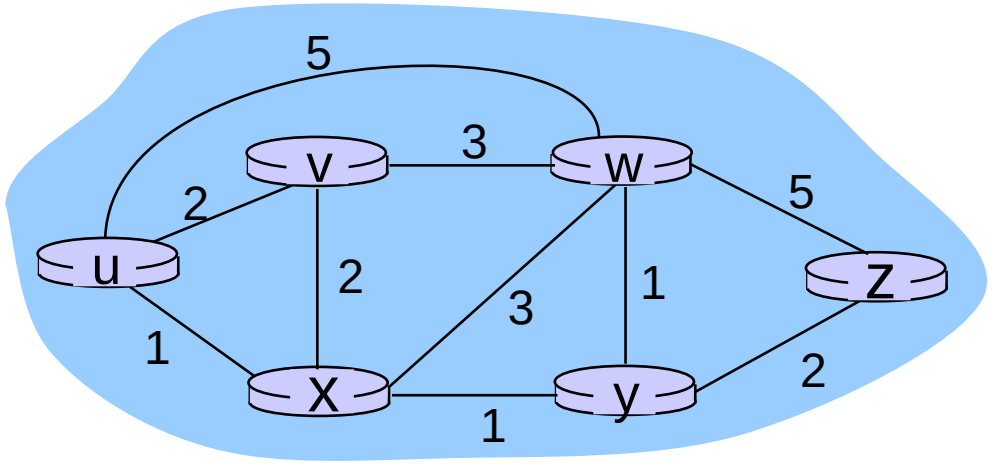
- run routing algorithms/protocol (RIP, OSPF, BGP)
- *forwarding* datagrams from incoming to outgoing link

Control Plane = routing
Vs
Data Plane = forwarding



SubnetNumber	SubnetMask	NextHop
128.96.34.0	255.255.255.128	Interface 0
128.96.34.128	255.255.255.128	Interface 1
128.96.33.0	255.255.255.0	R2

Graph abstraction



graph: G = (N,E)

N = set of routers = { u, v, w, x, y, z }

E = set of links ={ (u,v), (u,x), (v,x), (v,w), (x,w), (x,y), (w,y), (w,z), (y,z) }

X→ Z

Cost (x,v,w,z) = cost(x,v) + cost (v, w) + cost(w,z) = 10

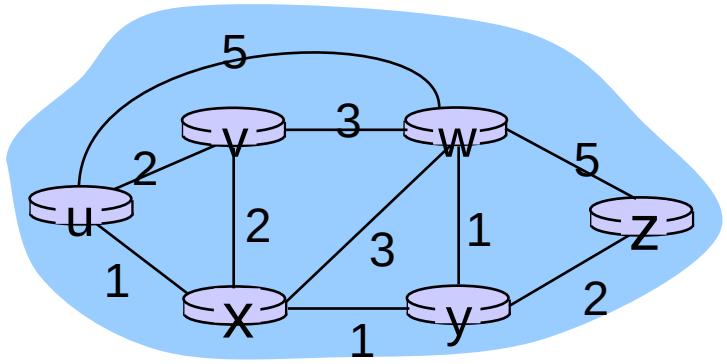
Cost (x,w,z) = cost(x,w) + cost(w,z) = 8

Cost(x, y, z) = ?

Objective → find the lowest cost path between all nodes

Dijkstra's Shortest-Path Algorithm

- Given a graph (network) with link costs
- Find the lowest cost paths to all nodes

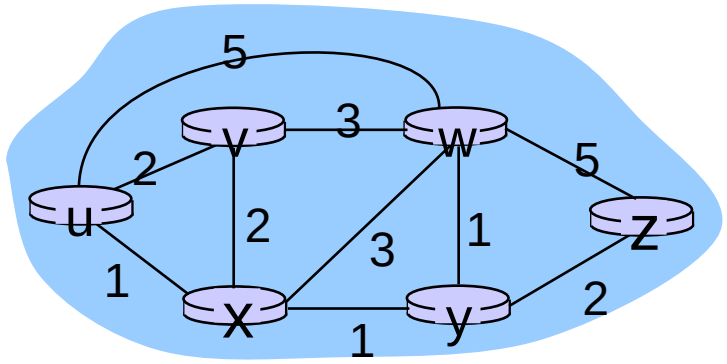


- Iterative
  - After n iterations, you will find least cost path to *n nodes*

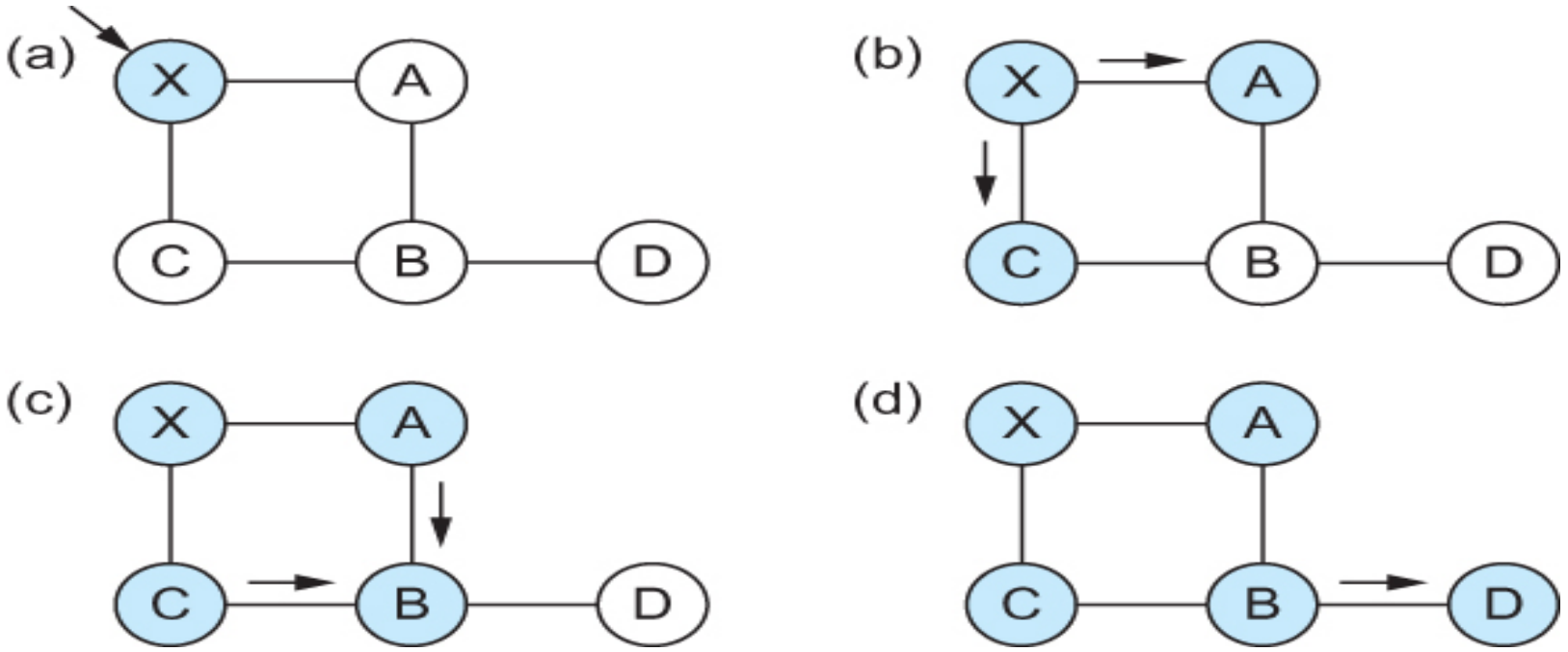
- *S = Least cost paths already known, initially source node {U}*
- *D(v): current cost of path from source(U) to node V*
  - *Initially, D(v) = c(u,v) for all nodes v adjacent to u*
  - *D(v) = ∞ for all other nodes*
  - *Update D(v) as we go*

Dijsktra's → Link State Routing

- Each node keeps track of adjacent links
- Each router broadcasts it's state
- Each router runs Dijkstra's algorithm
- Each router has complete picture of the network
- Example: Open Shortest Path First (OSPF)



Link State Routing – controlled flooding

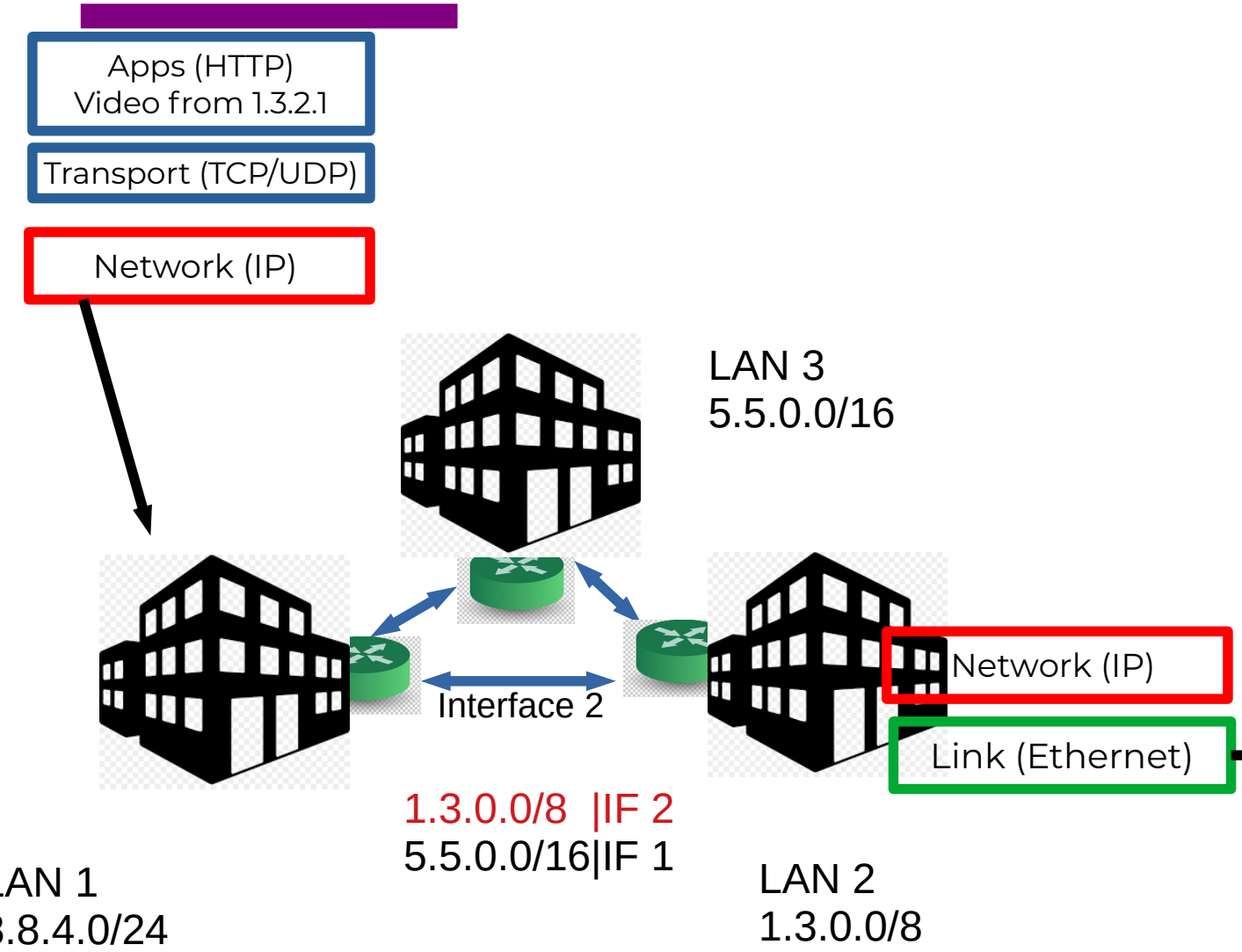


Flooding of link-state packets. (a) LSP arrives at node X; (b) X floods LSP to A and C; (c) A and C flood LSP to B (but not X); (d) flooding is complete

Link State Routing – controlled flooding

- Flood when topology changes or link goes down
  - Detected by periodic hello messages
  - If message missed → link down
- Refresh and flood periodically
- Problems?
  - High computational cost
  - Reliable flooding may not be reliable

Routing – Summarized

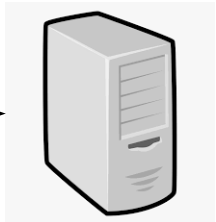


Routing will get you to the door (to another network)

A routing table tells you the most efficient way to get there

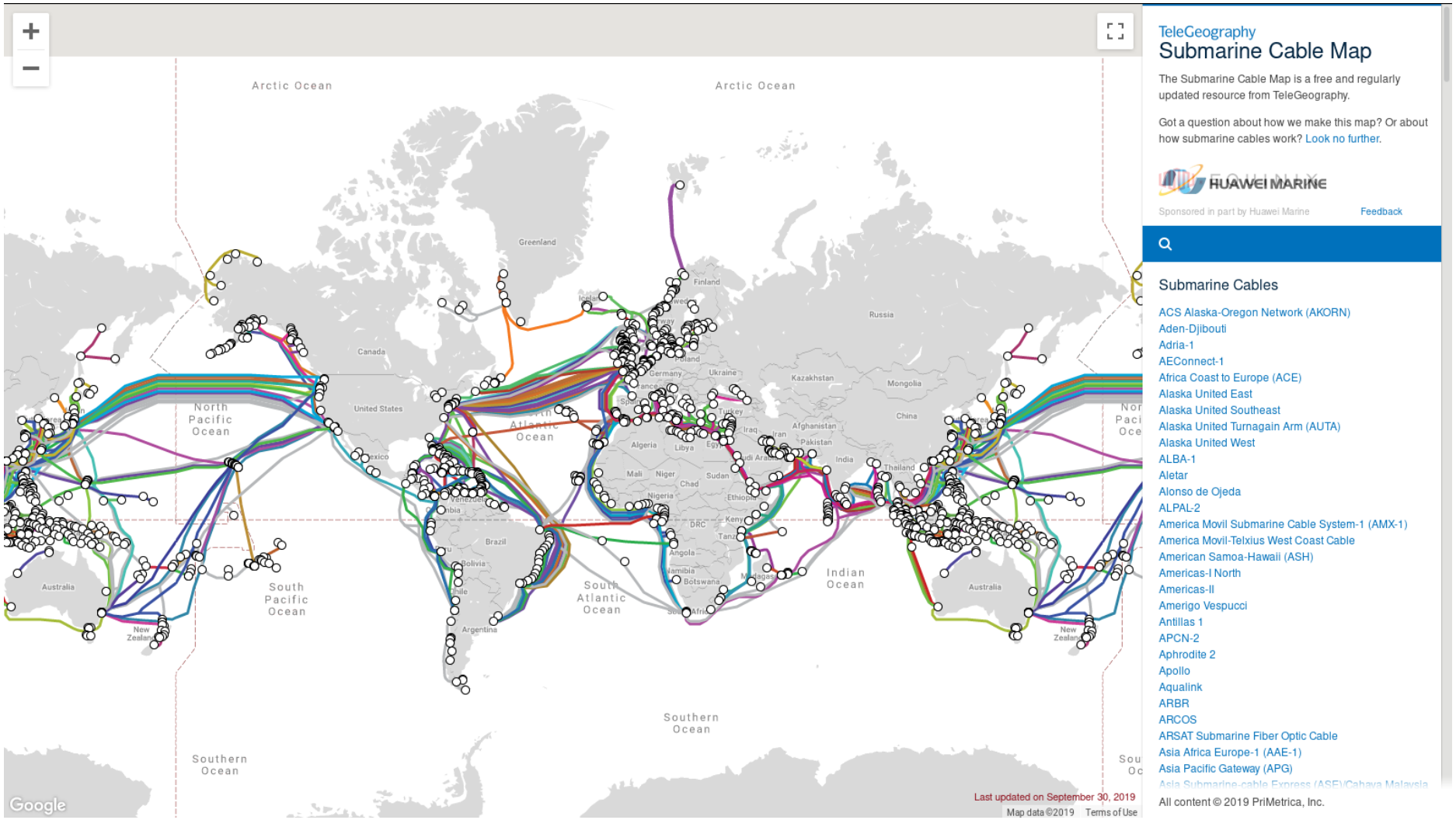
Once inside the building, use Layer 3 to Layer 2 mapping get to the actual hosts

IP: 1.3.2.1 → MAC:52:54:00:86:38:14

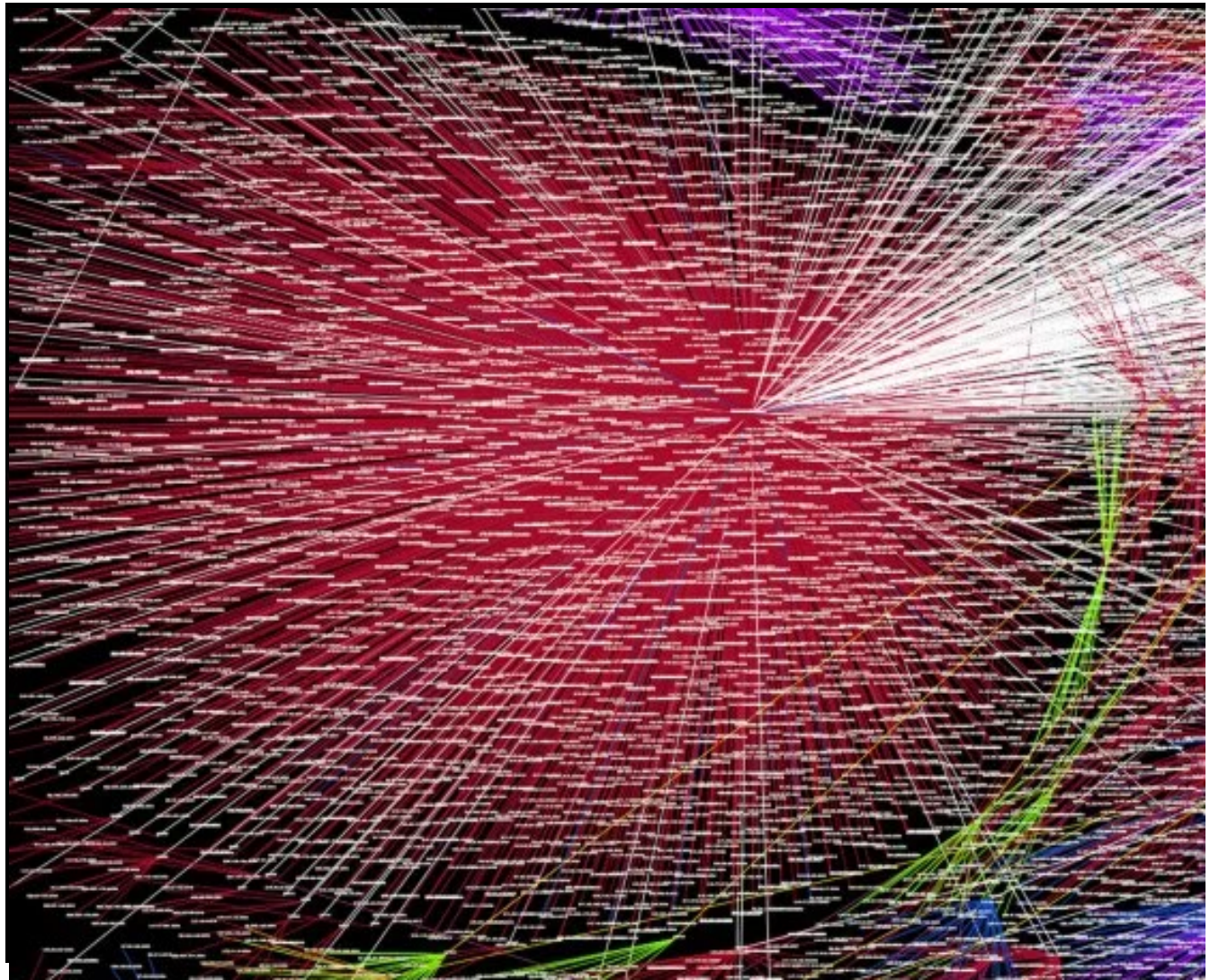


How do we scale this thing?

https://
www.submarinecablemap.com/



How do we scale this thing?



2003

2006

2009

2012

2015

http://www.opte.org/

https://time.com/3952373/internet-opte-proje

Local Routing – Gets you to the door.

What gets you to the campus?

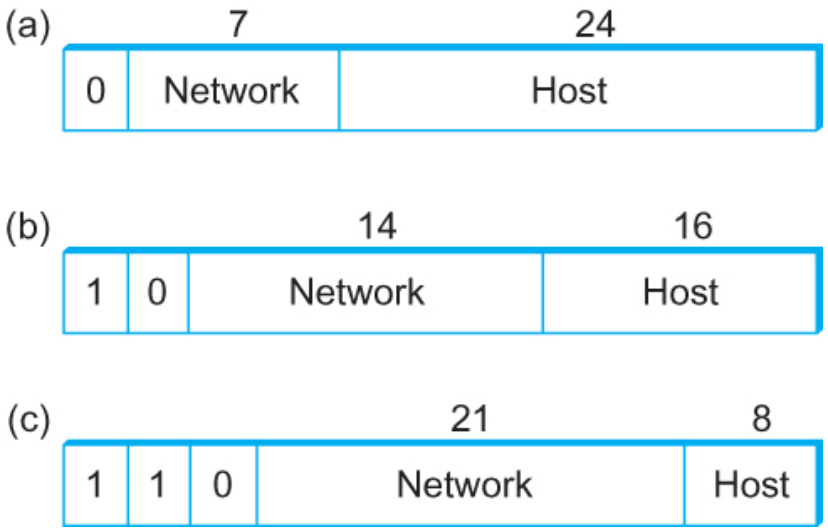
Interdomain Routing

- How is it different than routing?

Before that - Back to Addressing

- A 32 bit number in quad-dot notation
- Identifies an ***Interface***
  - ***A host might have several interfaces!!!***
- *129.82.138.254*

  10000001.01010010.10001010.11111110

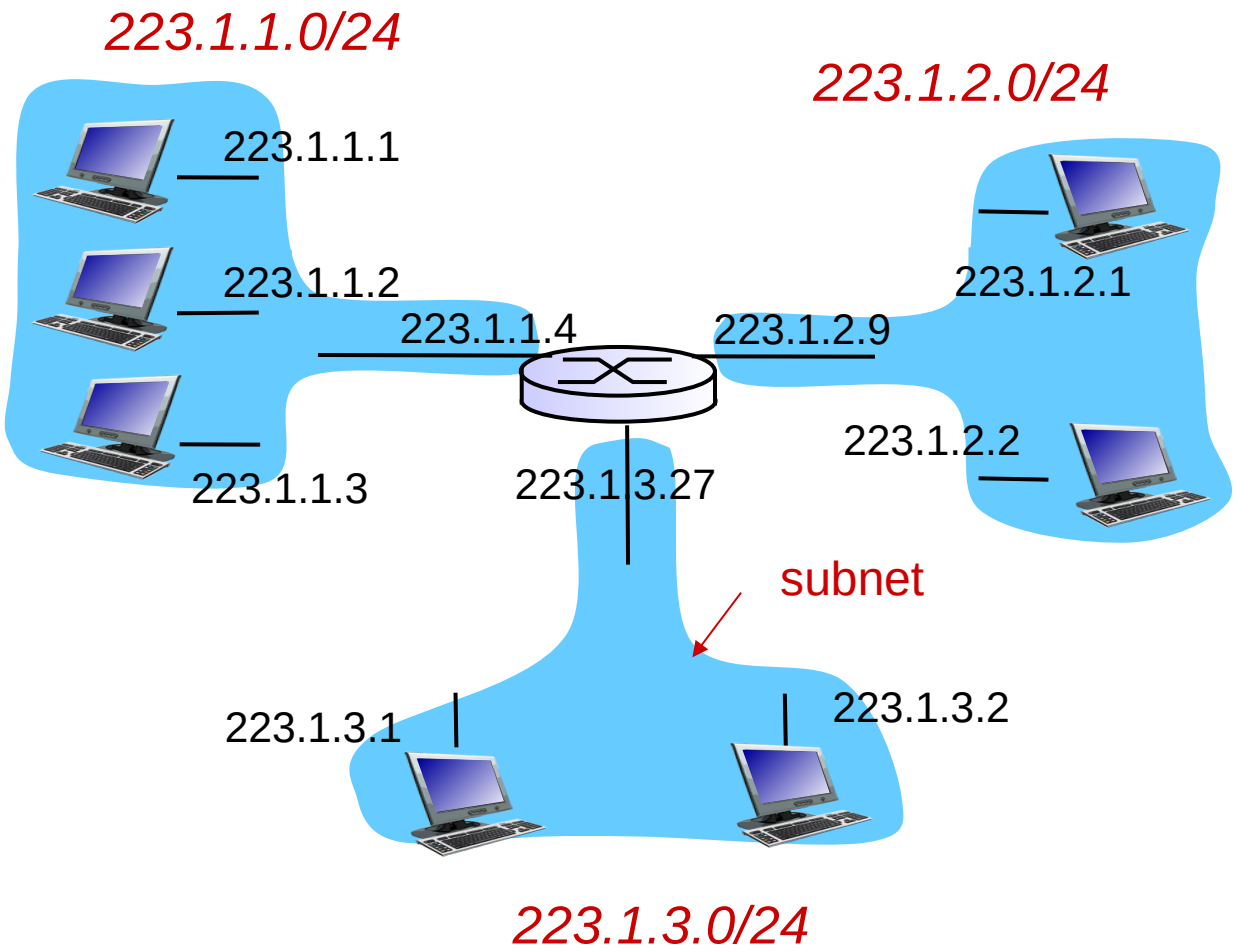


Subnets Revisited

Recipe:

- Create isolated networks – *subnets*

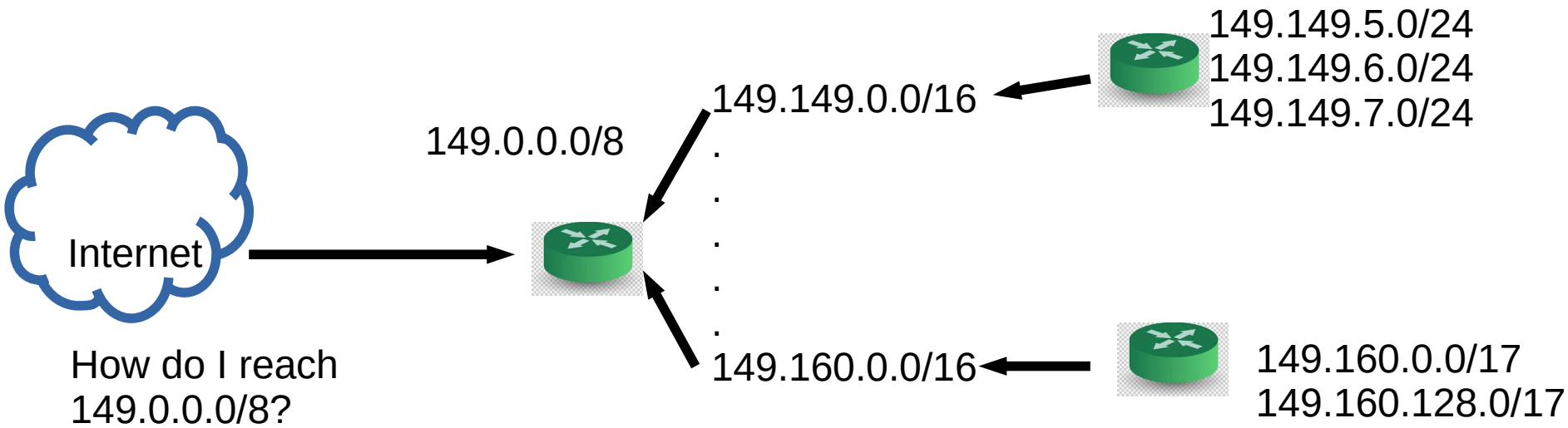
- No longer need to know individual Ips – knowing the subnet is enough
  - *223.1.1.0/14 → Interface 2*



subnet mask: /24

Subnets (Prefixes) scales the Internet

- Addresses are allocated in contiguous prefixes (tntech 149.149.0.0/16)
- Routing protocols operate based on prefixes (how do I reach 149.149.0.0/16)?

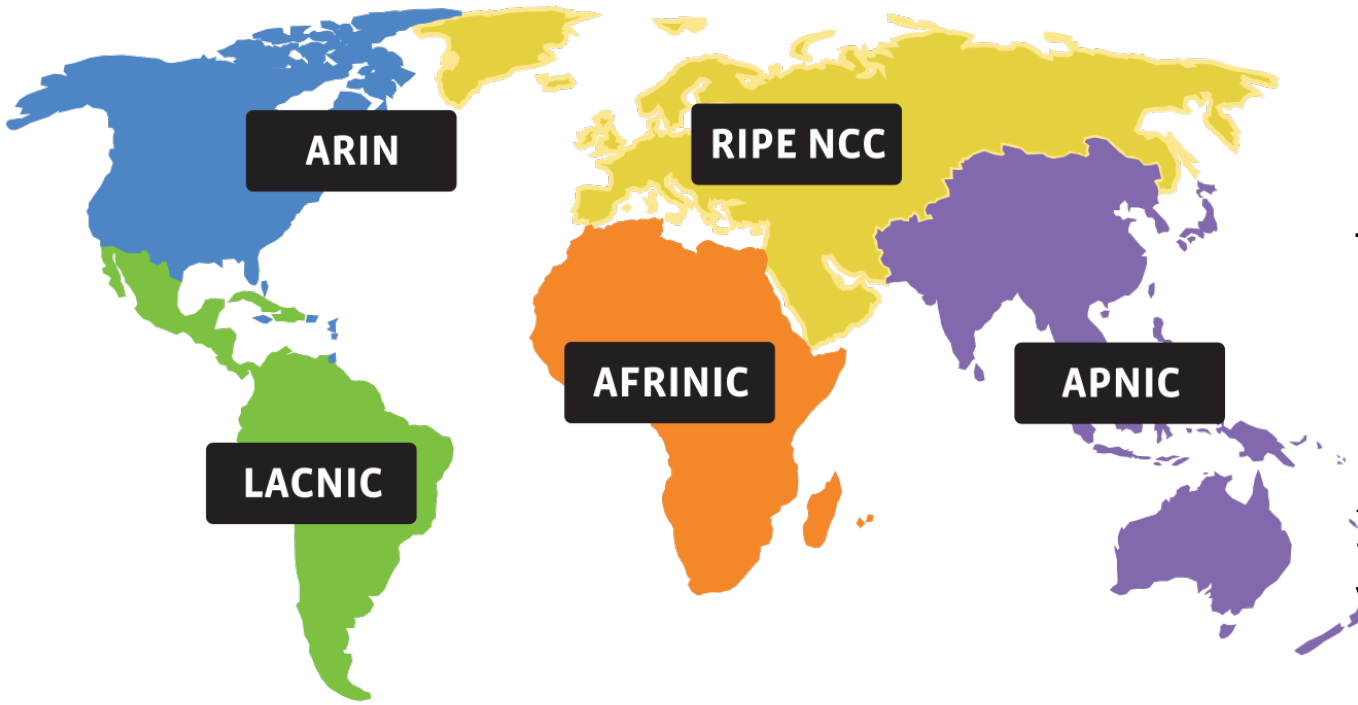


Not

How do I reach 149.149.5.0/24

How do I reach 149.149.6.0/24

Who gets what prefix?



0. Internet Corporation for Assigned Names and Numbers (ICANN) – Decides which RIRs get what address

1. Regional Internet Registries (RIRs) – Which orgs get what address

2. ISPs – Which customers get which address

How do you know who has a prefix?

whois

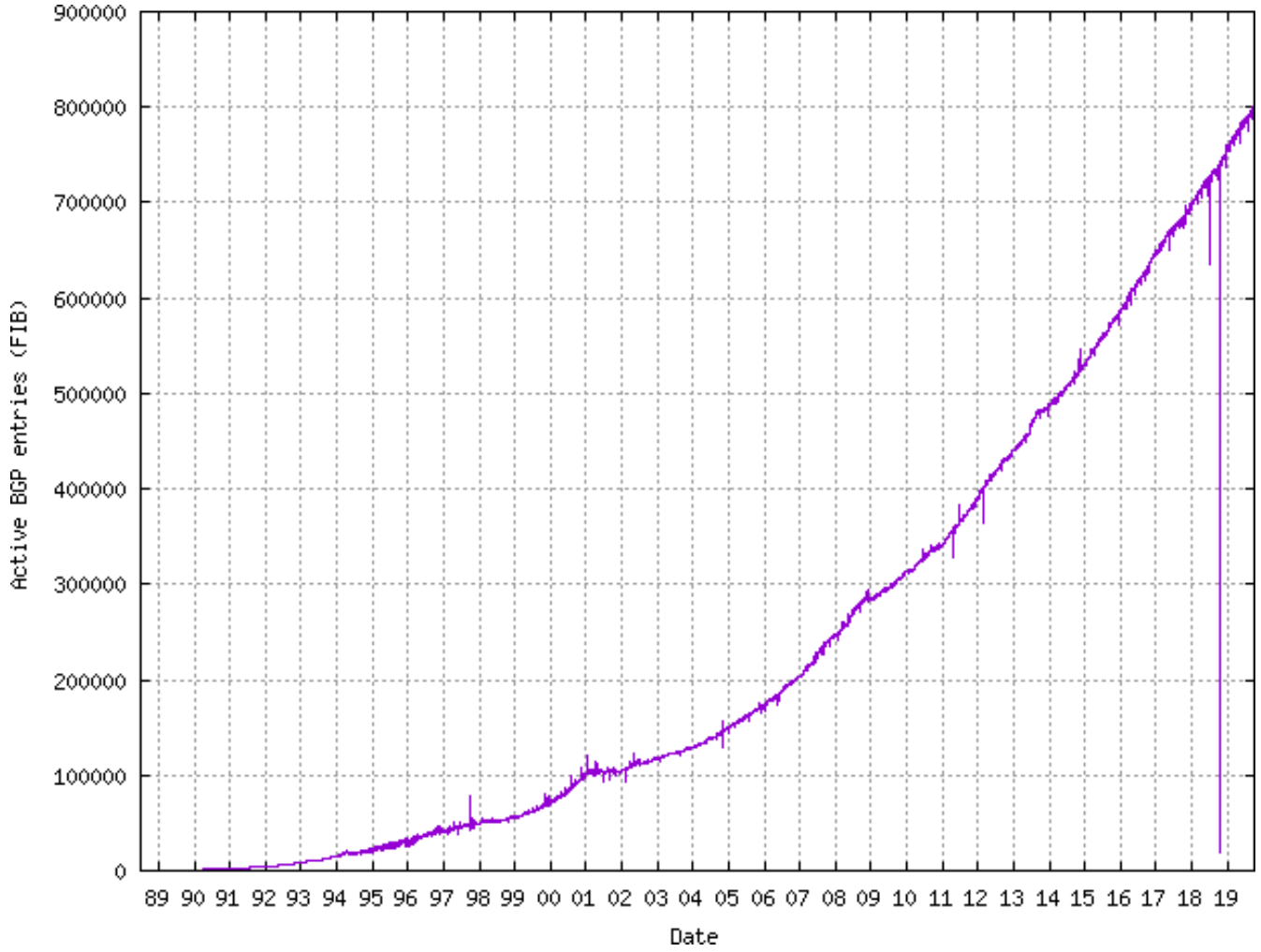
whois tntech.edu

Domain Name: TNTECH.EDU

Registrant:
Tennessee Technological University
Information Technology Service
1010 N. Peachtree Street
Cookeville, TN 38505
USA

Domain record activated: 09-Sep-1992
Domain record last updated: 26-Sep-2019
Domain expires: 31-Jul-2020

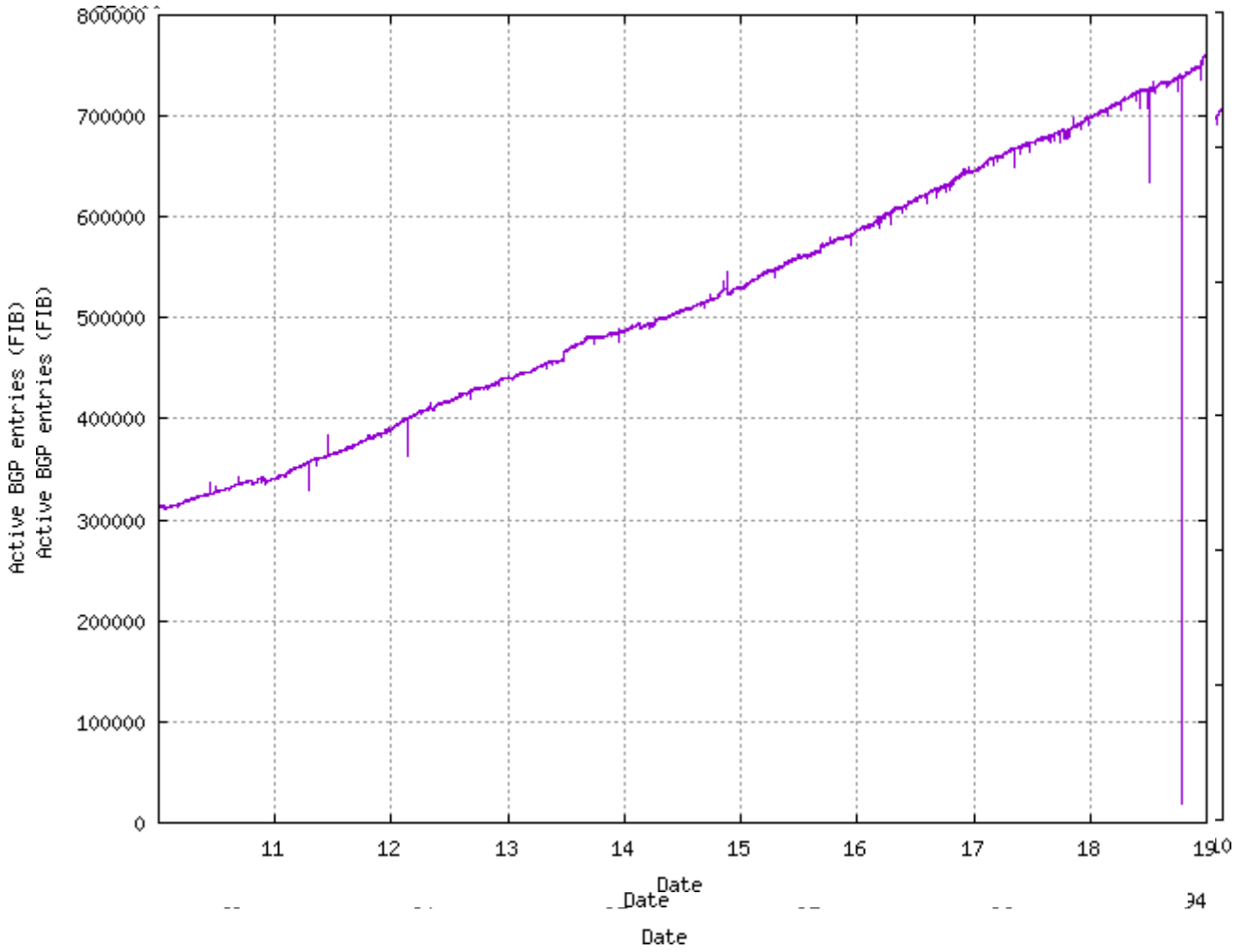
How many prefixes are there?



https://www.cidr-report.org/

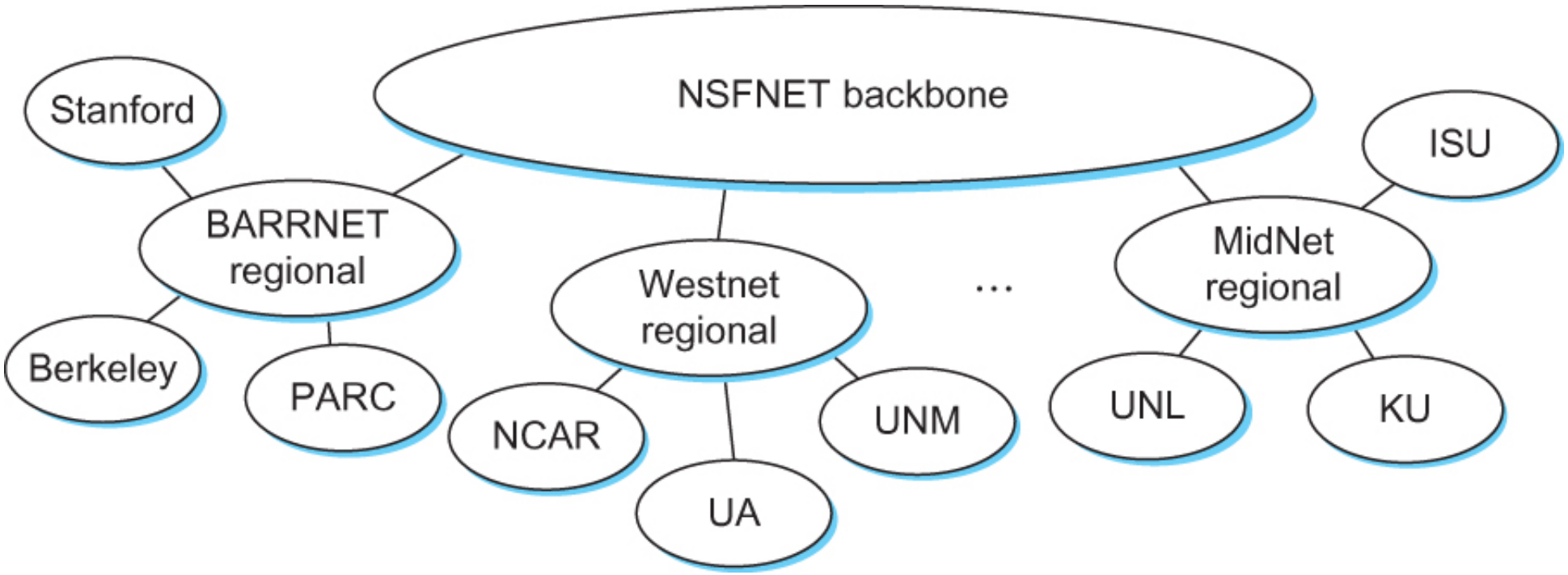
100K in 2001 → 800K in 2019

Bit of history – how the Internet evolved

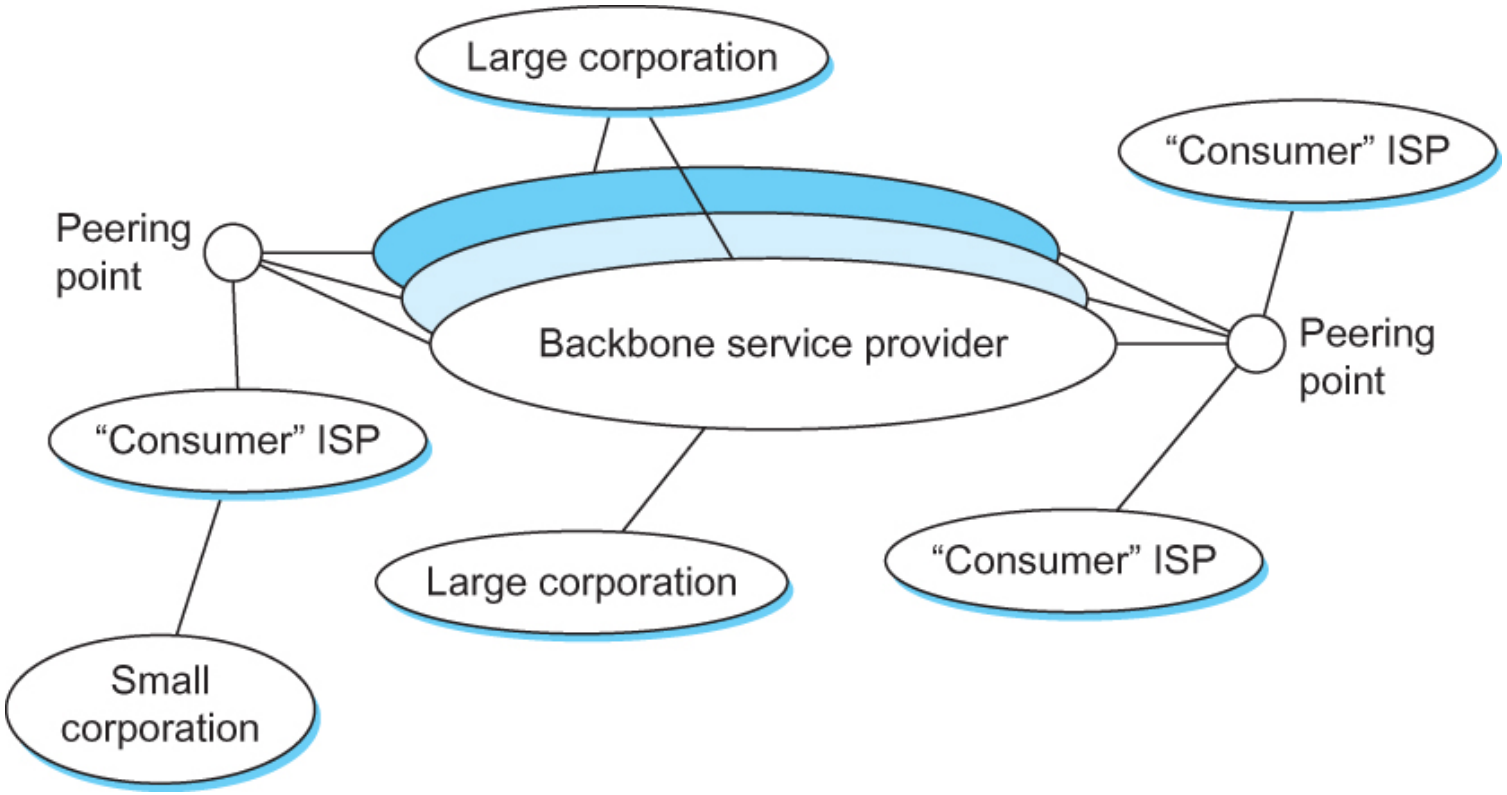


‘88-’94 - 0 → 14000
‘94-’00 – 90000 – Linear growth
’00-’10 – up to 300,000
’10-’19 – up to 800,000

Internet in the 1990s



Internet now



Interdomain routing - Policy

scale: with 600 million destinations and 800K networks:

- can't store all dest's in routing tables!
- routing table exchange would swamp links!

administrative autonomy

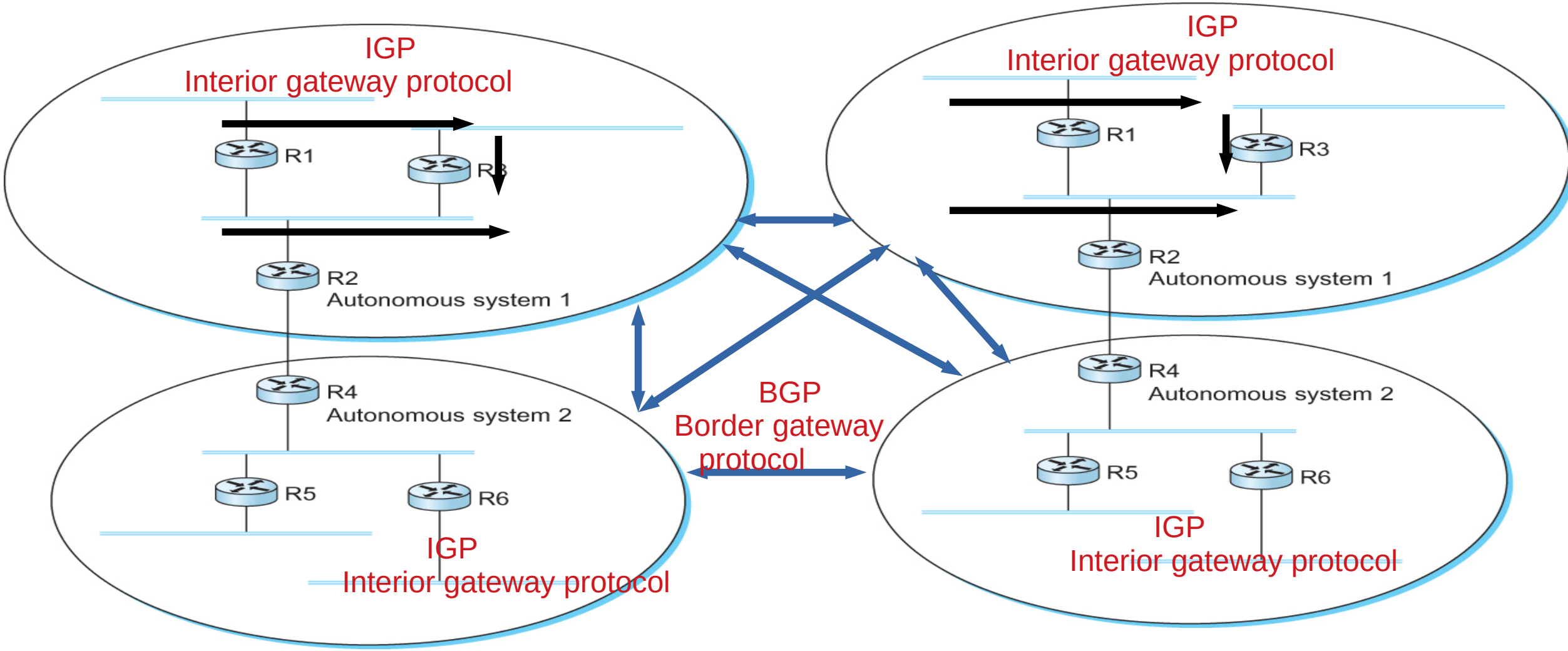
- internet = network of networks
- each network admin may want to control routing in its own network

Autonomous systems (ASes)

- AS
  - A set of routers under a single technical administration
  - Uses IGP within the AS to route packets
  - Uses BGP between Ases to route packets

- What happens inside an AS stays within that AS!
  - That is, AS decides routing metrics internally

Interdomain Routing



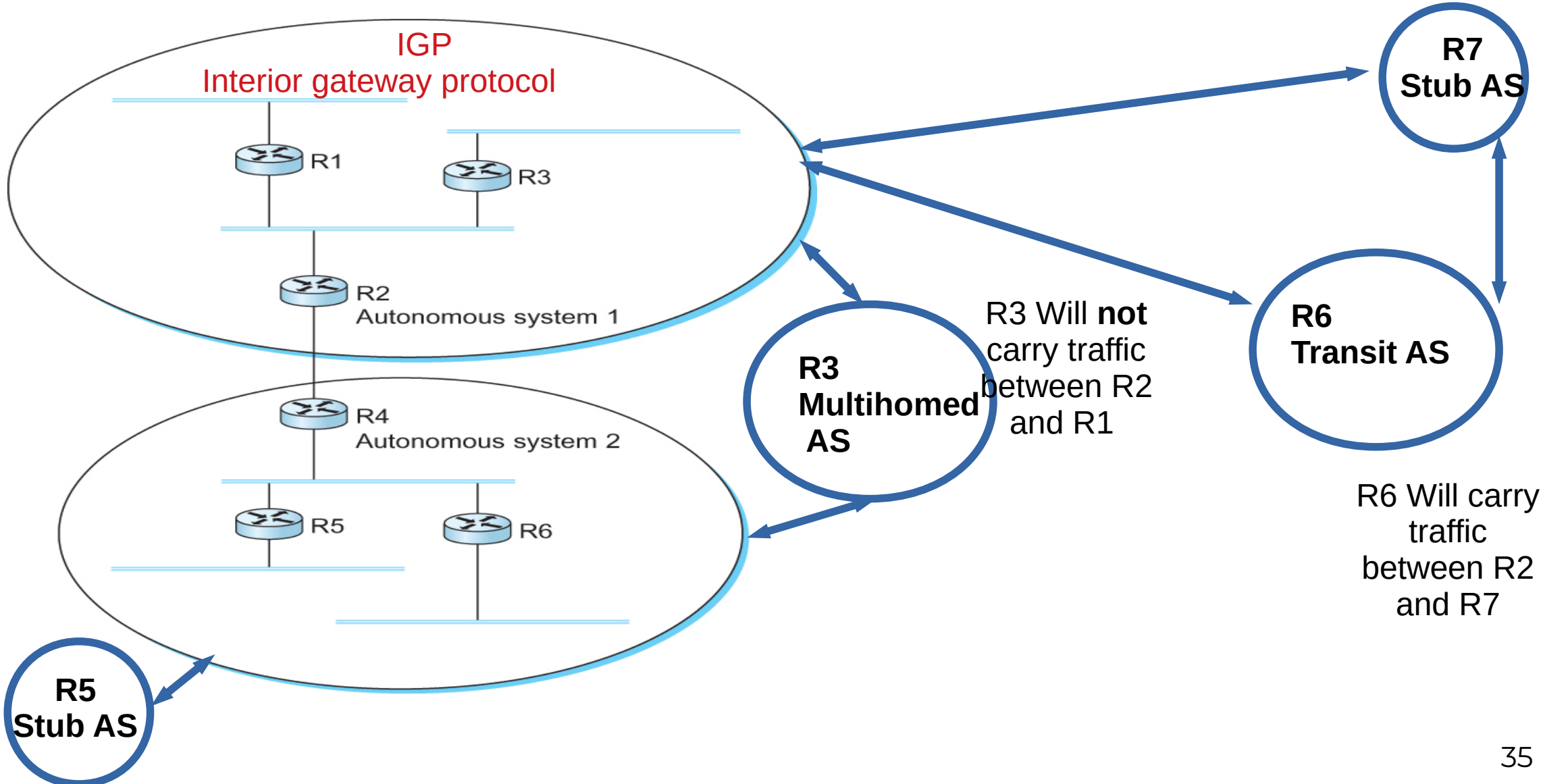
A network with four autonomous systems

BGP-4: Border Gateway Protocol

- Assumes the Internet is an arbitrarily interconnected set of AS's.

- Local traffic – within the AS
- Transit traffic – from AS1 to AS3 via AS2
- Three types of AS's
  - *Stub AS*
  - *Multihomed AS*
  - *Transit AS*

BGP-4: Border Gateway Protocol



BGP: Which routing protocol?

Link state?

- Does not scale
- you can have loops
- exposes routing costs to others

Distance vector?

- Slow to converge, count-to-infinity
- No universal metrics

BGP - goals

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

BGP - Goals

- Scalability: 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
  - impossible to calculate meaningful costs for a path crossing multiple ASs
  - A cost of 1000 is great at provider 1, terrible at provider 2
- Issues of trust
  - Provider A might be unwilling to believe certain advertisements from provider B

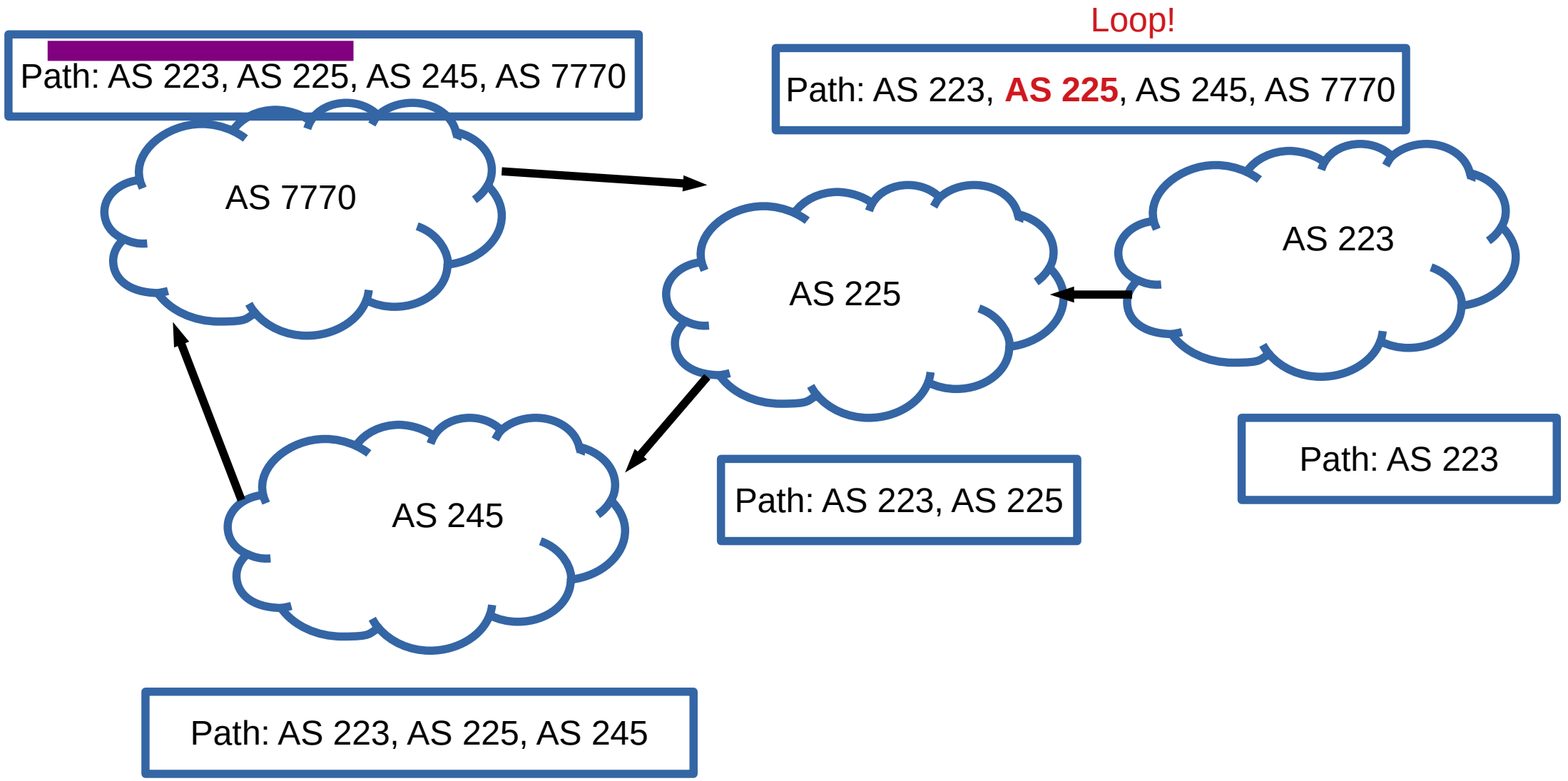
BGP: Path vector protocol

- Send the whole path with the routing update
- Loops are detected if an AS finds itself in the path
  - Reject if so
  - Accept otherwise
- Add self to the path and advertise to the neighbors
- Advantage: No loops, Local decision before advertizing

BGP: Path vector protocol

- Send the whole path with the routing update
- Loops are detected if an AS finds itself in the path
  - Reject if so
  - Accept otherwise
- Add self to the path and advertise to the neighbors
- Advantage: No loops, Local decision before advertising

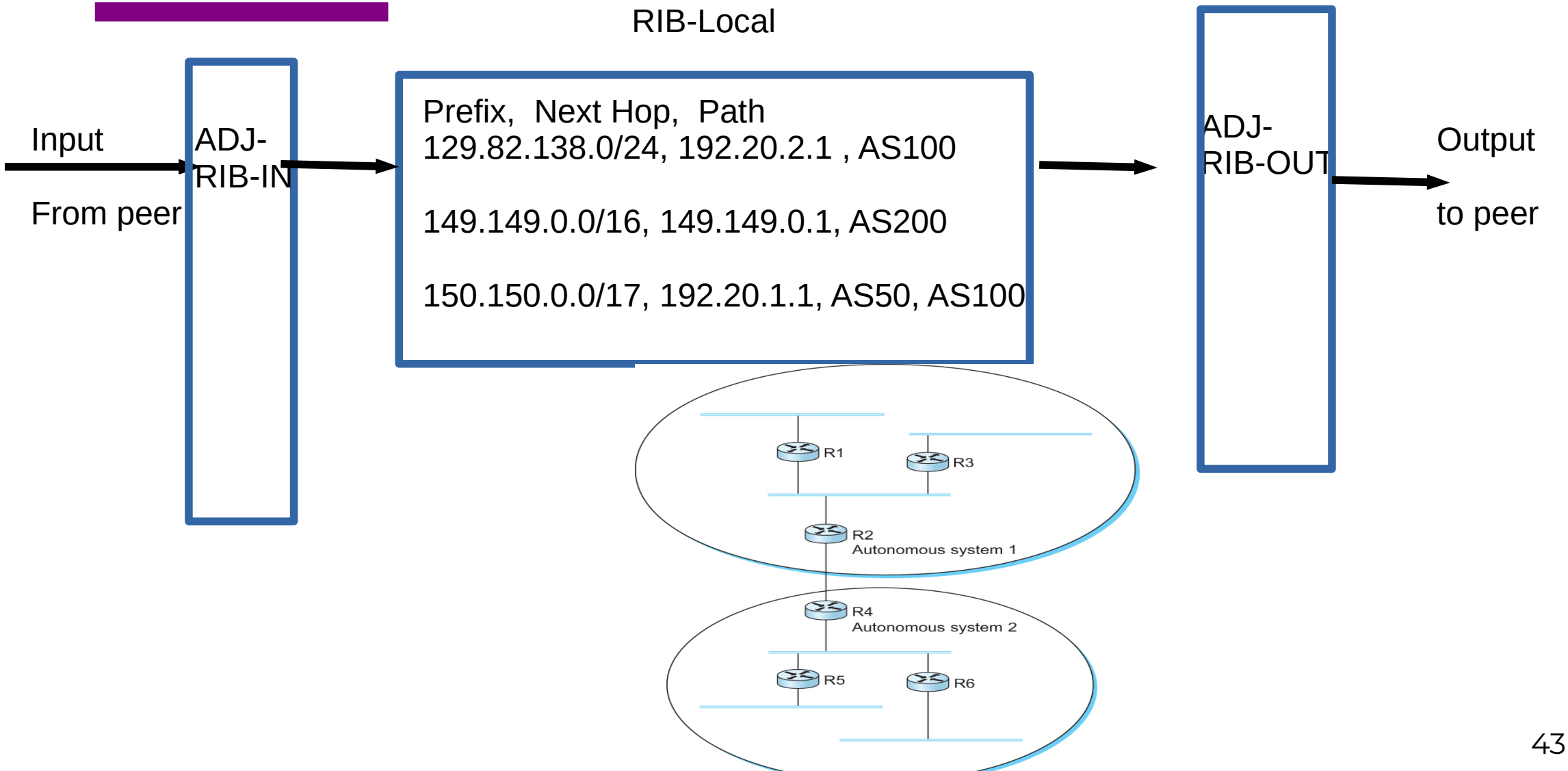
BGP: Path vector protocol



BGP: Interconnections

- Uses TCP port 179 to connect to **peers**
- Arbitrary connections between AS's
- Advantages:
  - Much simpler, no periodic update
  - Valid as long as TCP connection is valid (or withdrawn)
  - Incremental update (only a portion of the routing table)
- Disadvantages:
  - No security
  - Congestion control on routing messages

Routing Information Bases (RIB)



BGP: Hop by Hop model

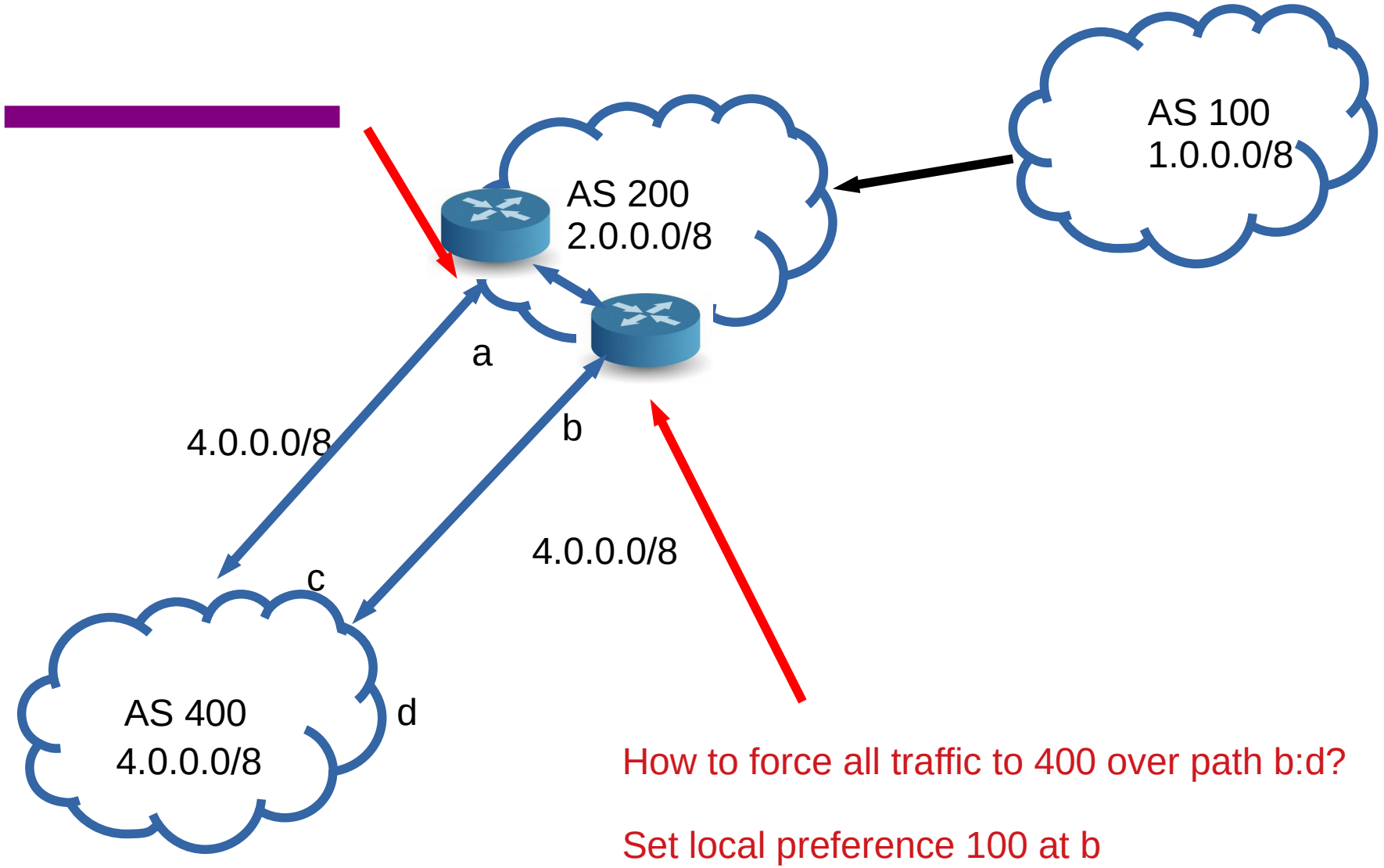
- You can only tell others what you are using
  - But you control what you say

- BGP advertises only to peers
  - Tell them what you are using
  - Hop-by-hop model

BGP: Allows for policy

- Capable of enforcing various policies
  - AS2 → Don’t use AS1 to get to AS3
- Not part of BGP – configuration information that controls propagation of paths

BGP Attributes – LOCAL-PREF Example



How to force all traffic to 400 over path b:d?

Set local preference 100 at b
Set local preference 200 at a

BGP Decision process

At ADJ-RIB-IN calculate degree of preference until one route for each destination remains!!

- select route with highest LOCAL-PREF
- Select route with shortest AS-PATH
- Select route with lowest MED
- Select route with smallest NEXT-HOP cost
- Select route learned from E-BGP peer with lowest ID
- Select route learned from I-BGP peer with lowest ID

- Install selected route in LOC-RIB

- Update ADJ-RIB-OUT, notify peers
  - You can only send what is in LOC-RIB (or a subset of it)

Switching vs Routing vs Interdomain Routing

- What do you think?
  - Switching is within a network
  - Routing is between networks – tells you how to get there
  - Interdomain routing – finds *a path* to all the networks

Conclusions

- Is BGP optimal?
  - No – Labovitz paper

- If we have a huge number of networks
  - everyone wants content that are spread around the world

  - How do we place content near the user?
    - Routing? BGP does not give you optimality
    - How about we trick the network?
      - Anycast – RFC 1546