Computer Networks

L8 – Network Layer III

Lecturer: CUI Lin

Department of Computer Science
Jinan University

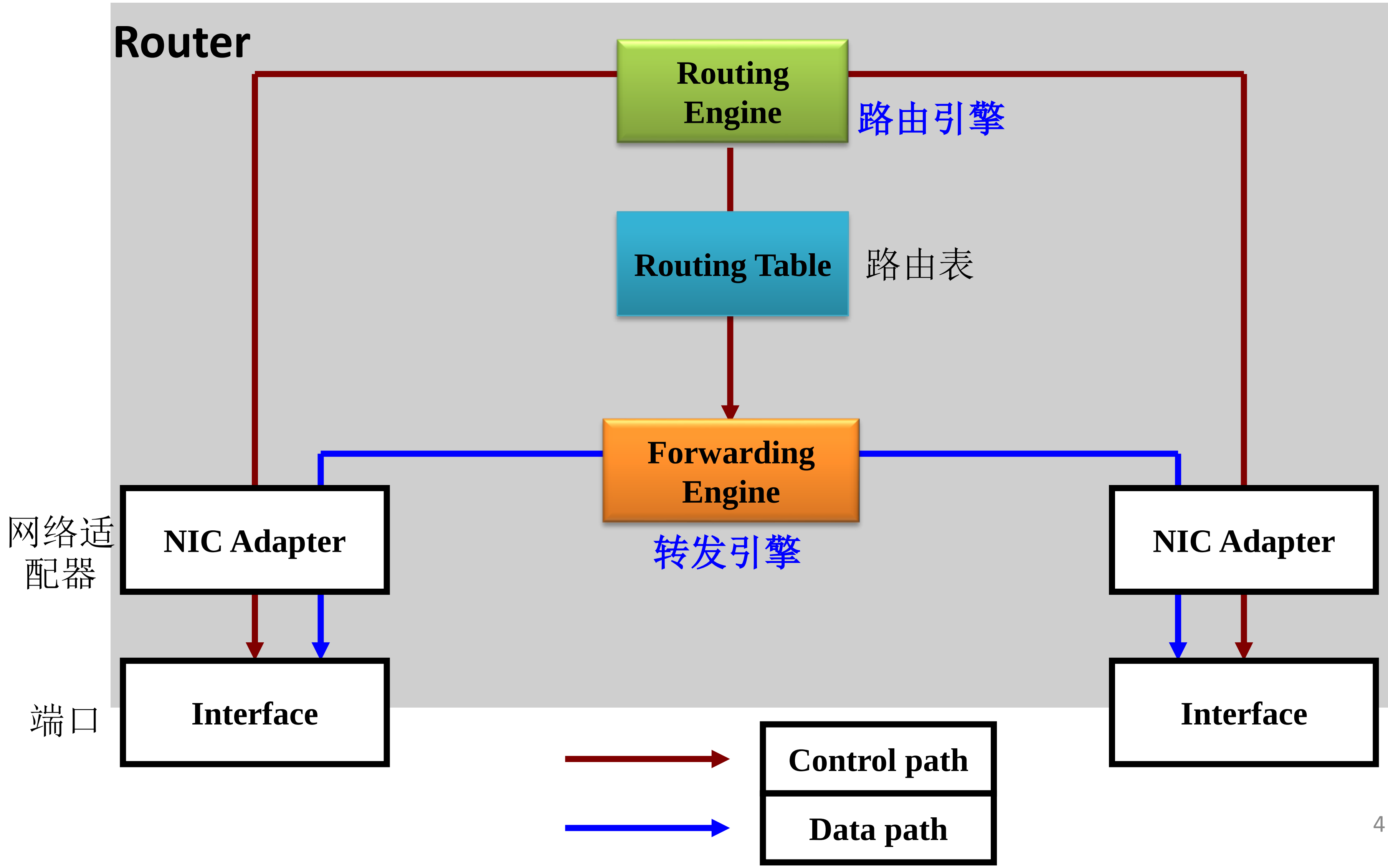
Topics in Network Layer

- Design Issues
- Internetworking
- Network Layer in the Internet
- Routing Algorithms
- Internet Routing and Multicasting

Routing Algorithms

- The optimality principle
- Shortest path algorithm
- Flooding
- Distance vector routing
- Link state routing
- Hierarchical routing
- Broadcast routing
- Multicast routing
- Anycast routing

Logical Architecture of a Router/L3 Switch



Routing versus Forwarding (1)

A router as having two processes inside:

- Forwarding process (转发): handles each packet as it arrives, looking up the outgoing line (or next hop) in the routing tables.
- Routing process(路由): responsible for filling and updating the routing tables
  - this is where the routing algorithm comes into play

Routing versus Forwarding (2)

- *Routing* is the process by which all nodes exchange control messages to calculate the *routes* (路由/路径) that packets will follow
  - Distributed process with *global* goals; Its emphasis is *correctness*
  - Nodes build a routing table that models the global network
- *Forwarding* is the process by which a node examines packets and sends them along their *paths* through the network
  - Involves *local* decisions; emphasis is *efficiency*
  - Nodes obtain *next hop* from their routing table

Routing Algorithms

- Routing is the process of discovering network paths
  - Model the network as a graph of nodes and links
  - Decide what to optimize (e.g., fairness vs. efficiency)
  - Update routes for changes in topology (e.g., failures)
- Datagrams networks: routing decision must be made anew for every arriving data packet since the best route may have changed since last time
- Virtual circuits networks: routing decision is made only when a new virtual circuit is being set up

Types of Routing Algorithms

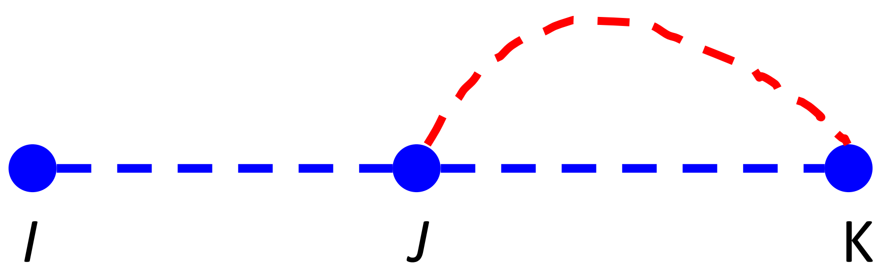
- Non-adaptive algorithms / Static routing
  - The routing table is computed in advance, off-line, and downloaded to the routers when the network is booted
  - Routes never change once initial routes have been selected
  - Because it does not respond to failures, static routing is mostly useful for situations in which the routing choice is clear
- Adaptive algorithms / Dynamic routing
  - Use such dynamic information as current traffic, topology, delay, etc. to select routes
  - Change their routing decisions to reflect changes in the topology, and usually the traffic as well

Delivery models

- Unicast:
  - A packet is sent to a single destination
- Broadcast:
  - A packet is sent to all destinations
- Multicast:
  - A packet is sent to a group of destinations
- Anycast:
  - A packet is sent to the nearest member of a group

The Optimality Principle (1)

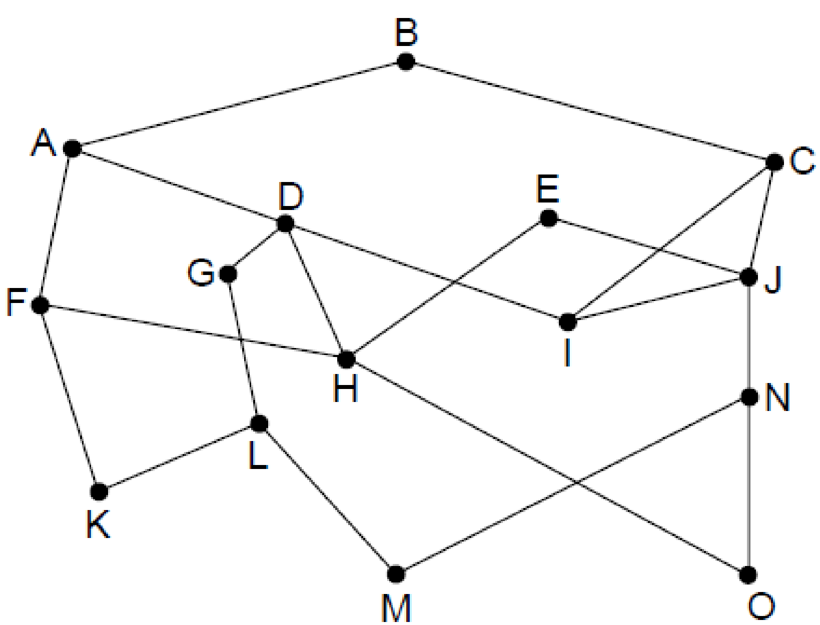
- This simply states that if router *J* is on the optimal path from router *I* to router *K*, then the optimal path from *J* to *K* also falls along this same path.



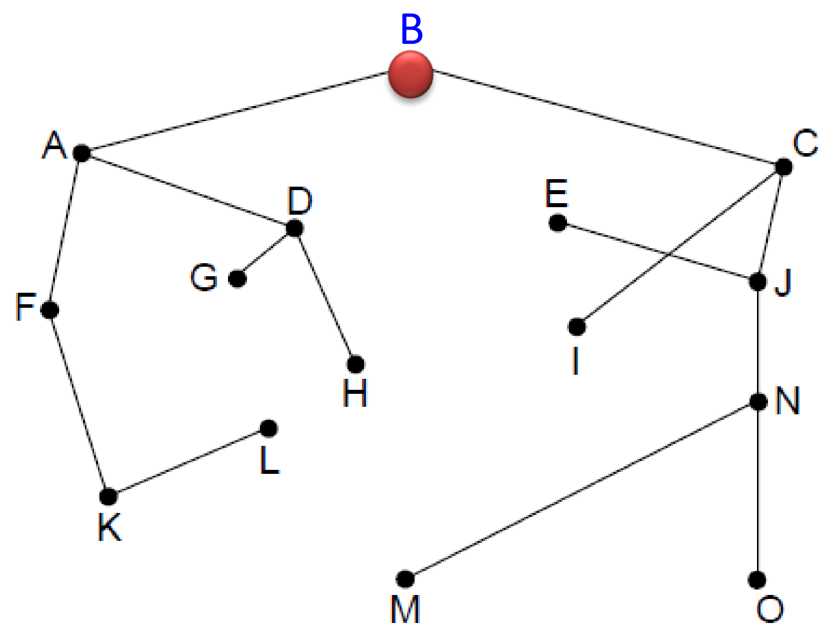
- Each portion of a best path is also a best path

The Optimality Principle (2)

- All optimal routes from a station to other stations in the network, jointly constitute a sink tree
- Routers have to collaborate to build the sink tree for each source station or destination station



Network Topology



Sink tree of best paths (hops) to router B

Graph abstraction

- Graph abstraction for routing algorithms:
  - Nodes are routers
  - Edges are physical links
  - link cost/weight: delay, cost, bandwidth, congestion level, etc.
- To choose a route between a given pair of routers: find "Good" path
  - Typically means minimum cost path
  - Other definitions possible

Shortest Path Algorithm

- Often used by routing algorithms because it is simple and easy to understand
- Shortest Path Metrics (Path Length)
  - Number of Hops
  - Physical Distance
  - Mean Queuing and Transmission Delay
  - Bandwidth
  - Average Traffic
  - Communication Cost

Shortest Path Algorithm

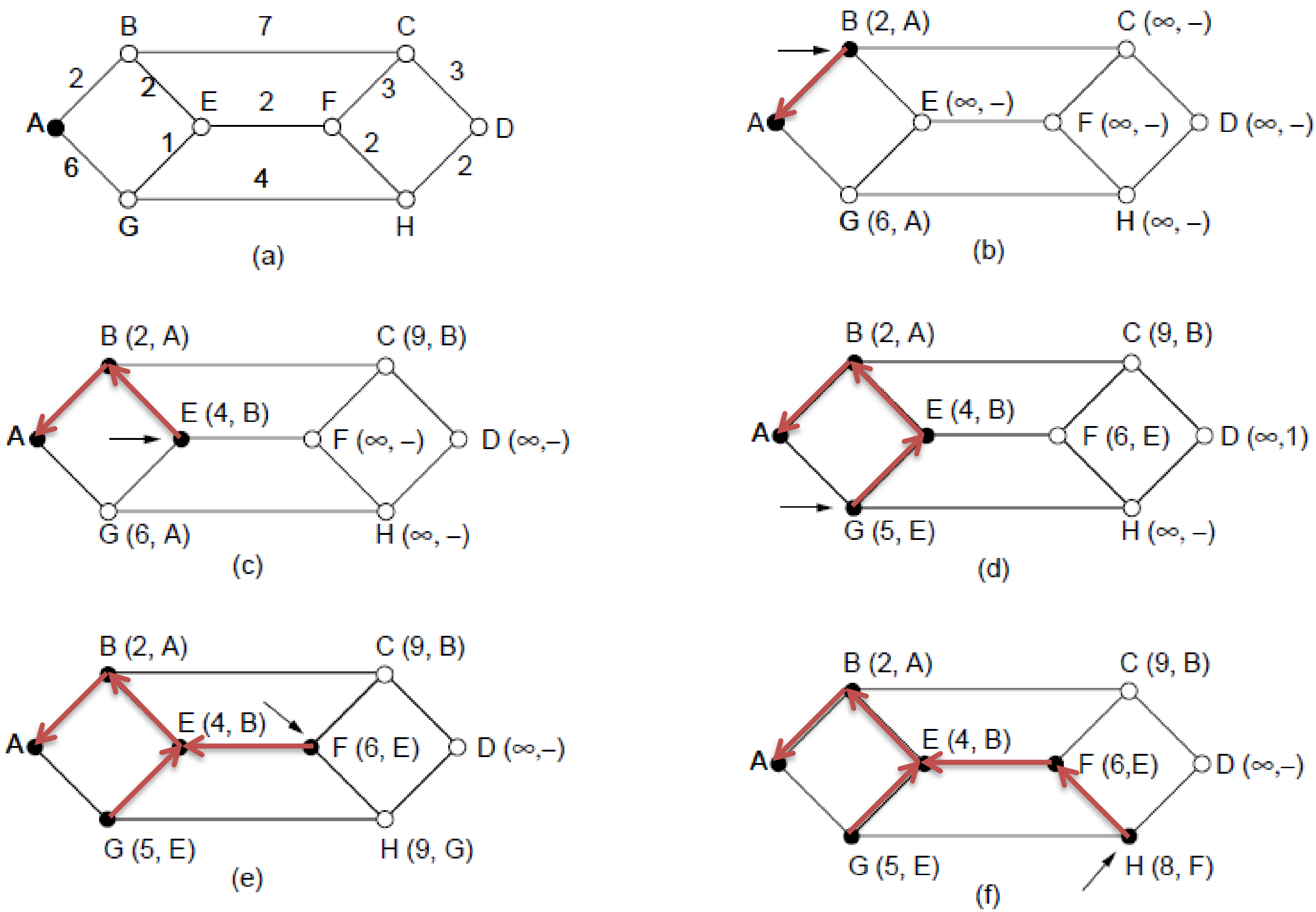
Dijkstra's algorithm computes a sink tree on the graph:

- Each link is assigned a non-negative weight/distance
- Shortest path is the one with lowest total weight
- Using weights of 1 gives paths with fewest hops

Algorithm:

- Start with sink, set distance at other nodes to infinity
- Relax distance to other nodes
- Pick the lowest distance node, add it to sink tree
- Repeat until all nodes are in the sink tree

Shortest Path Algorithm



A network and first six steps in computing the shortest paths from A to D. Red arrows show the sink tree so far.

Flooding (泛洪)

- A simple method to send a packet to all network nodes
- Basic idea: Forward an incoming packet across every outgoing line, except the one it came through
- Basic problem: how to avoid “drowning by packets”?
  - Use a hop counter: after a packet has been forwarded across *N* routers, it is discarded.
  - Be sure to forward a packet only once (i.e. avoid directed cycles).
    - Requires sequence numbers per source router.

Flooding – Reduce Looping

- Each router maintains a private sequence number. When it sends a new packet, it includes sequence number in the packet, and increase its sequence number.
- For each source router *S*, a router:
  - Keeps track of the highest sequence number seen from *S*
  - Whenever it receives a packet from *S* containing a sequence number lower or equal to the one stored in its table, it discards the packet
  - Otherwise, it updates the entry for *S* and forwards the packet

Flooding uses

- Flooding has several important uses:
  - An effective approach for broadcasting information to all nodes;
  - Several copies of the same packet may reach nodes
  - Tremendously robust: e.g., military applications
- Theoretical-chooses all possible paths, so it chooses the shortest one.

Distance Vector Routing

Important

- Distance Vector (DV, 距离矢量算法) is a distributed routing algorithm
  - Shortest path computation is split across nodes
  - Distributed version of Bellman-Ford algorithm, used in the Internet (ARPANET) under name RIP (Routing Information Protocol)
  - It works, but very slow convergence after failure

Distance Vector Algorithm

- Each router computes its routing table in a distributed setting:
  - Each router only knows the distance (cost) of to its neighbors (e.g., send echo requests)
  - Each router can talk only to its neighbors using messages
  - All routers run the same algorithm concurrently
  - Nodes and links may fail, messages may be lost

Distance Vector Algorithm

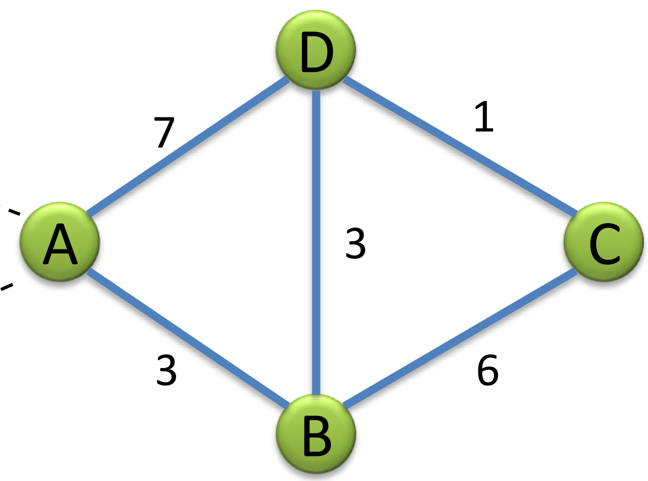
- Each router maintains a table (vector) of costs to all destinations as well as next hops
  - Initialize neighbors with known cost, others with infinity
  - Tables are updated by exchanging information with neighbors
- Routers periodically send copy of vector to neighbors
- On reception of a vector, if neighbor's path to a destination plus cost to that neighbor is better
  - Update the cost and next-hop in local table

- Assuming no changes, will converge to shortest paths

DV Algorithm Example

- Consider a simple network, each node run on its own
  - Node A can only talk to B and D

To	Cost
A	0
B	INF
C	INF
D	INF



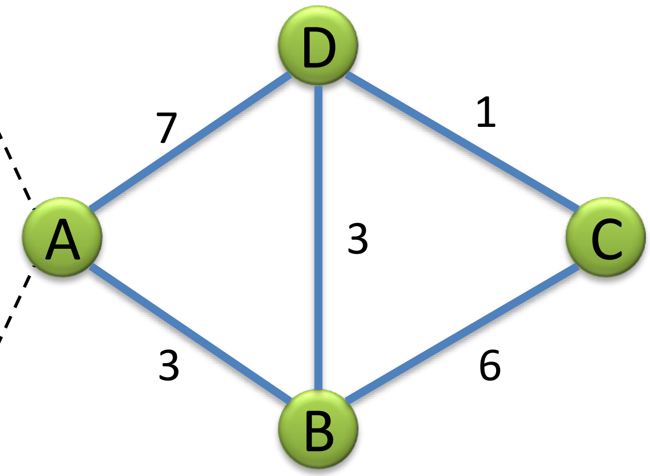
DV Algorithm Example

- First exchange, A hear from B and D, finds 1-hop routes
  – Node A always learns *min* (B+3, D+7)

To	B says	D says
A	INF	INF
B	0	INF
C	INF	INF
D	INF	0

B+3	D+7
INF	INF
3	INF
INF	INF
INF	7

To	Cost	Next
A	0	--
B	3	B
C	INF	--
D	7	D

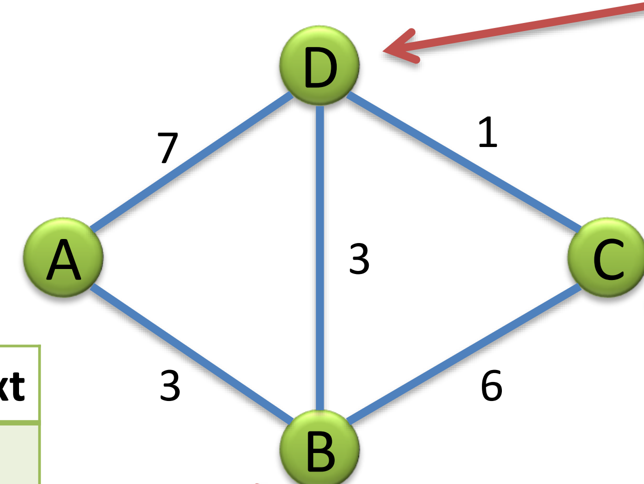


DV Algorithm Example

- Similarly, B, C and D can learn their 1-hop routes

To	Cost	Next
A	0	--
B	3	B
C	INF	--
D	7	D

To	Cost	Next
A	7	A
B	3	B
C	1	C
D	0	--



To	Cost	Next
A	3	A
B	0	--
C	6	C
D	3	D

To	Cost	Next
A	INF	--
B	6	B
C	0	--
D	1	D

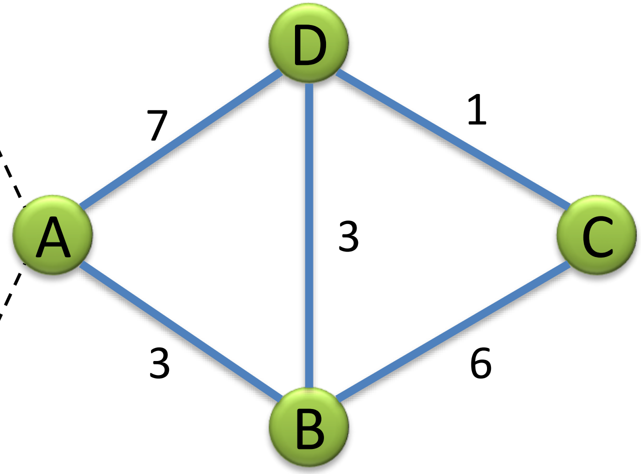
DV Algorithm Example

- Second exchange, A hear from B and D, finds 2-hop routes
  - Node A always learns min (B+3, D+7)

To	B says	D says
A	3	7
B	0	3
C	6	1
D	3	0

B+3	D+7
6	14
3	10
9	8
6	7

To	Cost	Next
A	0	--
B	3	B
C	8	D
D	6	B



This process continues to third and subsequent exchange, until converged…

The Count-to-Infinity Problem

- Failures can cause DV to "count to infinity" while seeking a path to an unreachable node

A	B	C	D	E	
	•	•	•	•	Initially
	1	•	•	•	After 1 exchange
	1	2	•	•	After 2 exchanges
	1	2	3	•	After 3 exchanges
	1	2	3	4	After 4 exchanges

Good news of a path to *A* spreads quickly

A	B	C	D	E	
	1	2	3	4	Initially
	3	2	3	4	After 1 exchange
	3	4	3	4	After 2 exchanges
	5	4	5	4	After 3 exchanges
	5	6	5	6	After 4 exchanges
	7	6	7	6	After 5 exchanges
	7	8	7	8	After 6 exchanges
		⋮			
	•	•	•	•	

Bad news of no path to *A* is learned slowly

B doesn't know whether itself is on the path from C

RIP (Routing Information Protocol)

- RIP is a DV protocol with hop count as metric
  - Infinity is 16 hops; limits network size
  - Measures to handle count-to-infinity: split horizon with poisoned reverse rule, ref. link
- Routers send vector every 30 secs
  - Run on top of UDP
  - Timeout in 180s to detect failures
- Specified in RFC 1058

Link State Routing (链路状态路由)

Also Important

- Link state is an alternative to distance vector
  - More computation but better dynamics
- Link State Routing (LSR) is widely used in practice
  - Used in Internet/ARPANET from 1979
  - Modern networks use OSPF and IS-IS

Link State Routing Algorithm

Proceeds in two phases:

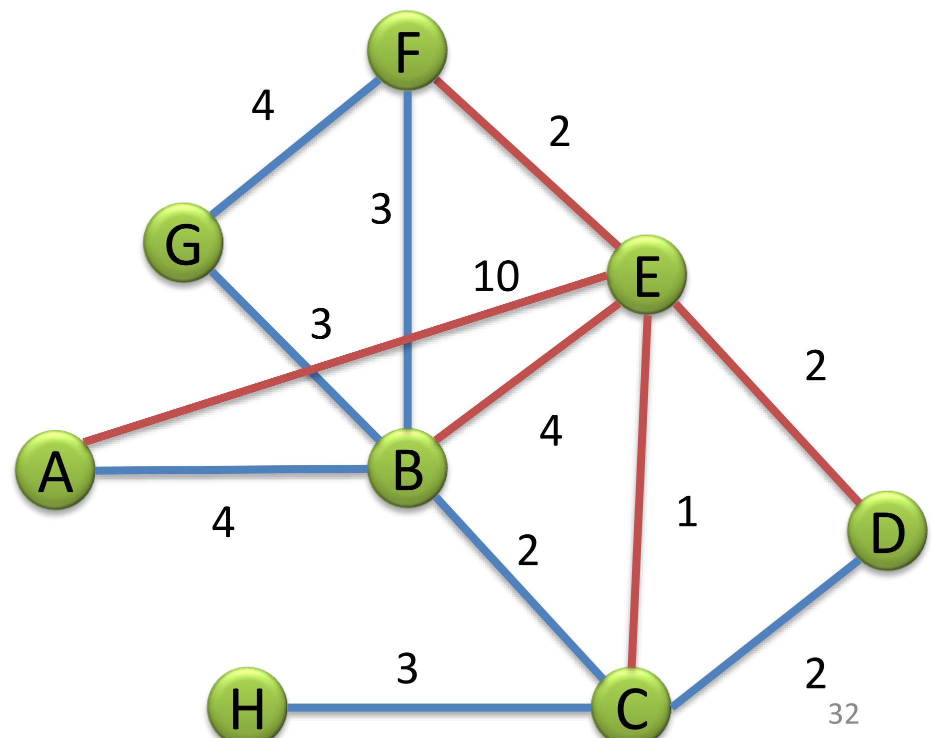
1. Each node floods information (topology) about its neighbors in LSPs (Link State Packets)
   – all nodes learn the full network graph
2. Each node runs Dijkstra’s algorithm (or equivalent) to compute the path to take for each destination

Phase 1: Topology Dissemination

- Each node floods LSP (Link State Packet)
  - Use HELLO msg to learn neighbors after boot up
  - LSP includes a list of neighbors and weights of links to reach them

Node E's LSP:

E	
Seq. No	
Age	
A	10
B	4
C	1
D	2
F	2



Reliable Flooding

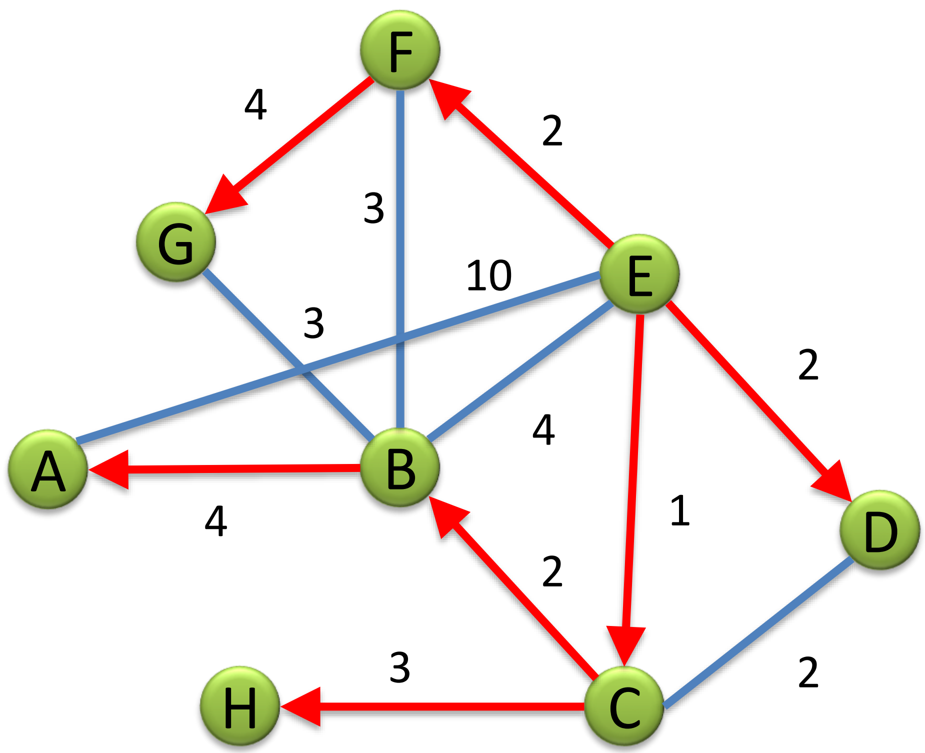
- Sequence number and Age are used for reliable flooding
  - 32 bit Sequence No.: 137 years to wrap around with one packet per second
  - Age to prevent Sequence number confusion: router crashes and lose track of its Seq., or the Seq. is corrupted
  - New LSPs are acknowledged on the lines when they are received and sent on all other lines

Phase 2: Route Computation

- Each node has full topology of entire network
  - By combining all LSPs
- Each node simply run Dijkstra's algorithm
  - Router will know which link to use to each destination
  - Compute the routing table

Routing Table Example

Source Tree for E (from Dijkstra):



E's Table:

To	Next
A	C
B	C
C	C
D	D
E	--
F	F
G	F
H	C

Goods and Bads of LSR

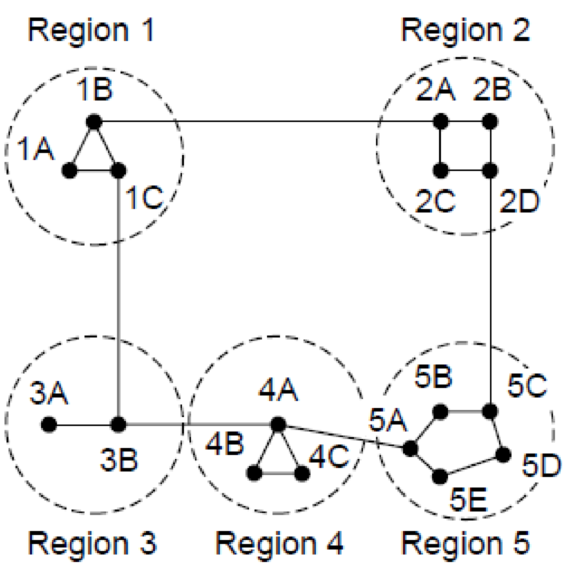
- Goods
  - Good consistency of each router information
  - Quick convergence for good and bad news
- Bads
  - Each router need large memory to store the input link states of other routers
  - The computation time can be an issue

Hierarchical Routing

- Problem: Scalability
  - Network grows in size, routing table grows proportionally
  - More router memory and CPU time are needed
- Go for *suboptimal* routes by introducing hierarchical routing and regions, and separate algorithms for intra-region and inter-region routing.

Hierarchical Routing

- Hierarchical routing reduces the work of route computation but may result in slightly longer paths than flat routing



Full table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2A	1B	2
2B	1B	3
2C	1B	3
2D	1B	4
3A	1C	3
3B	1C	2
4A	1C	3
4B	1C	4
4C	1C	4
5A	1C	4
5B	1C	5
5C	1B	5
5D	1C	6
5E	1C	5

Hierarchical table for 1A

Dest.	Line	Hops
1A	–	–
1B	1B	1
1C	1C	1
2	1B	2
3	1C	2
4	1C	3
5	1C	4

Best choice to reach nodes in Region 5 except for 5C

Broadcast Routing

- Broadcasting sends a packet to all destinations simultaneously
- Several ways to implement broadcasting:
  - Send a unicast packet to each destination separately
  - Flood packets to all nodes
  - Multi-destination routing:
    - Each packet contains a list (or bitmap) of all destinations
  - Use sink tree
  - etc.

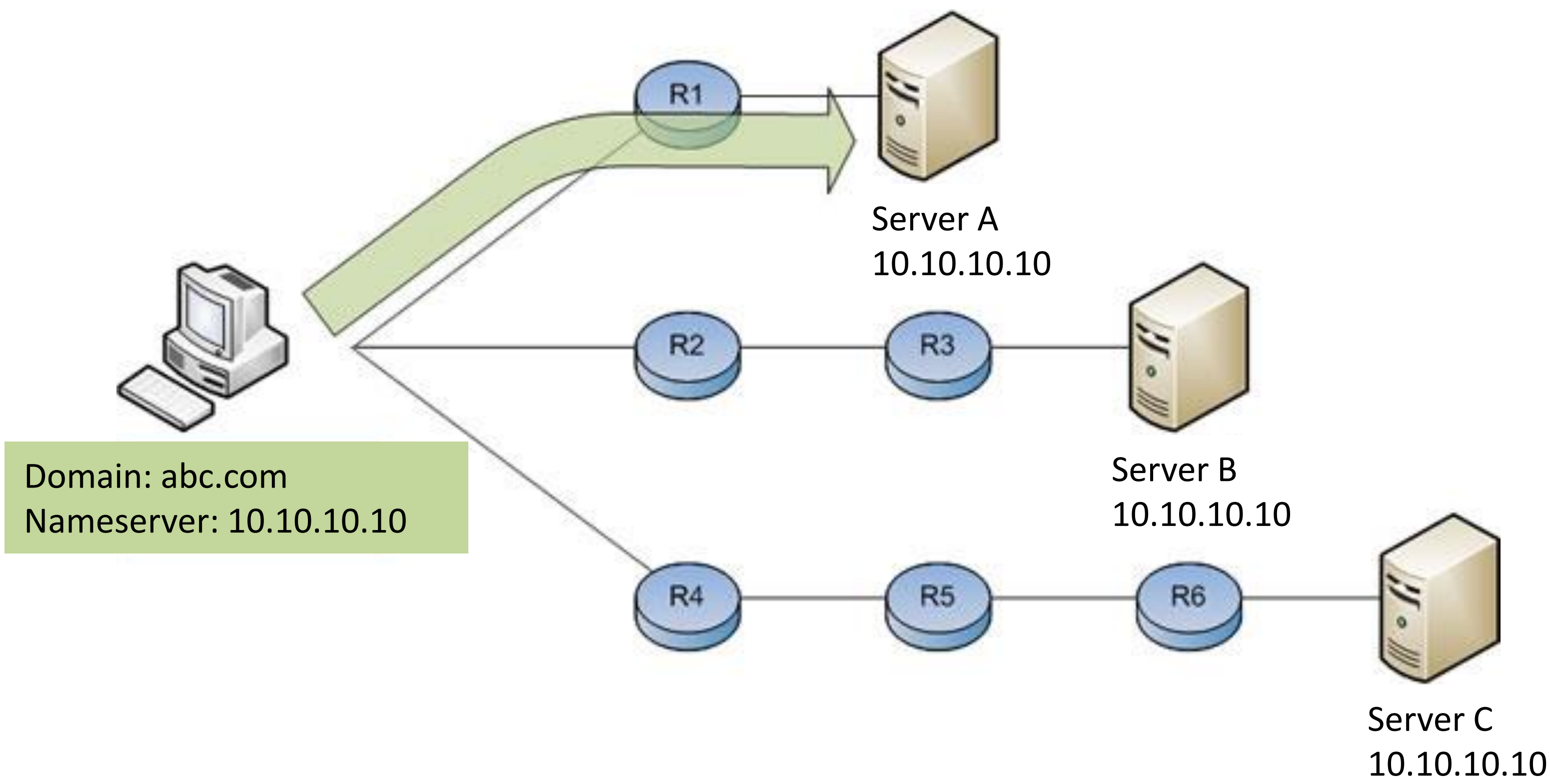
Multicast Routing

- Multicast sends message to a subset of the nodes, called group
- Examples: live video streaming
- Uses a different tree for each group and source
- Usually sending packets along a spanning tree

Anycast Routing

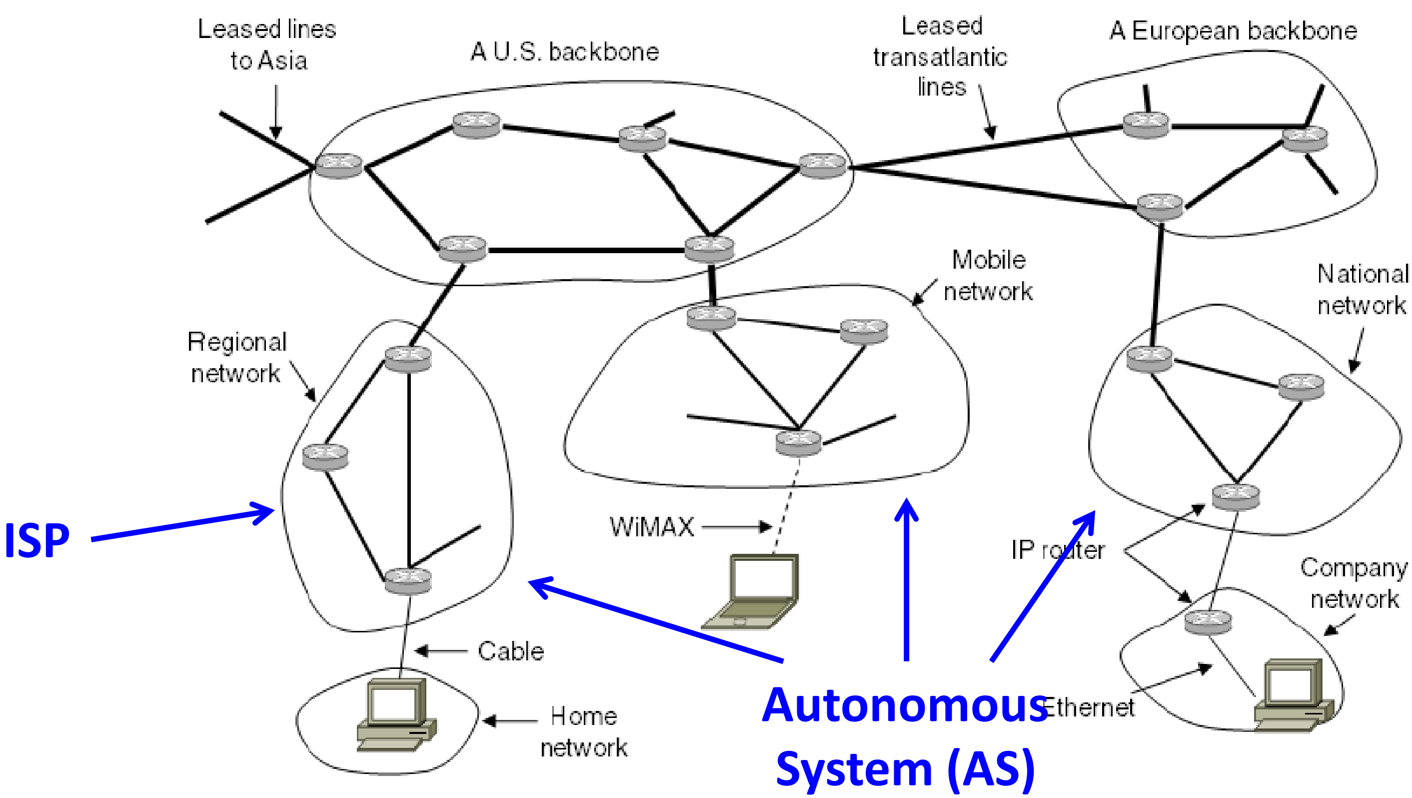
- For some service, it is getting the right information that matters, rather than the node that is contacted
  - E.g., NTP, CDN, DNS…
- Anycast Routing: a packet is delivered to the nearest member of a group
  - Can be designed based on distance vector and link state routing, by assigning one IP address to all servers in the group

Anycast Routing Example



Routing in the Internet

- The global Internet consists of Autonomous Systems (AS) interconnected with each other:
  - E.g., ISP, company, university



Routing in the Internet

Important

- The global Internet consists of Autonomous Systems (AS) interconnected with each other:
  - E.g., ISP, company, university
- Two-level routing:
  - Intra-AS/Intra-Domain:
    - Each network can use its own routing algorithm

ASrouting Alg

  - Inter-AS/Inter-Domain:

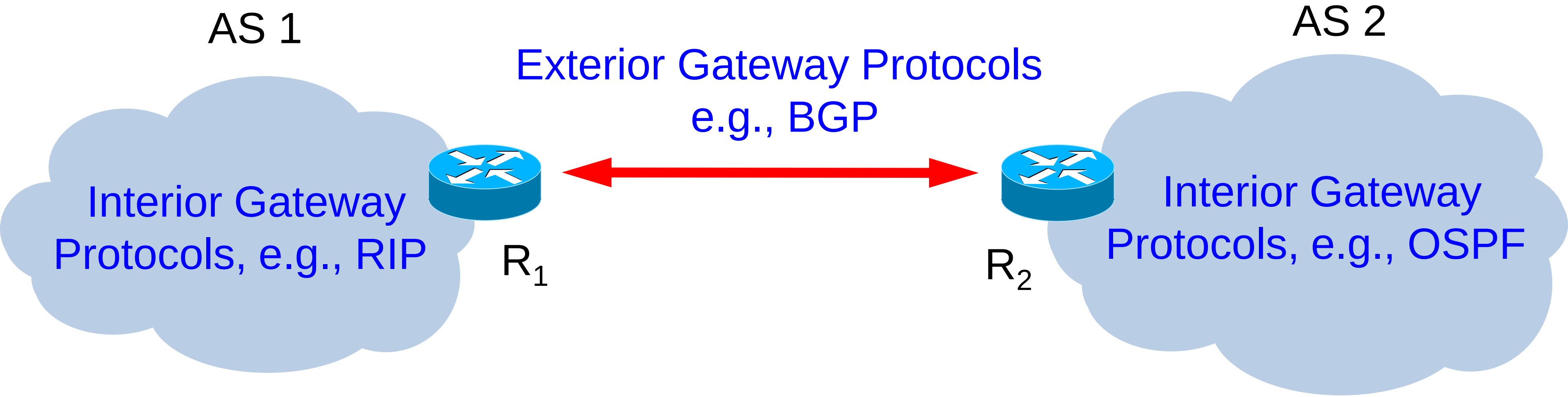
Routing Alg

    - All networks use same routing protocol
    - Networks may have different/conflicting goals

Need to use different protocols…

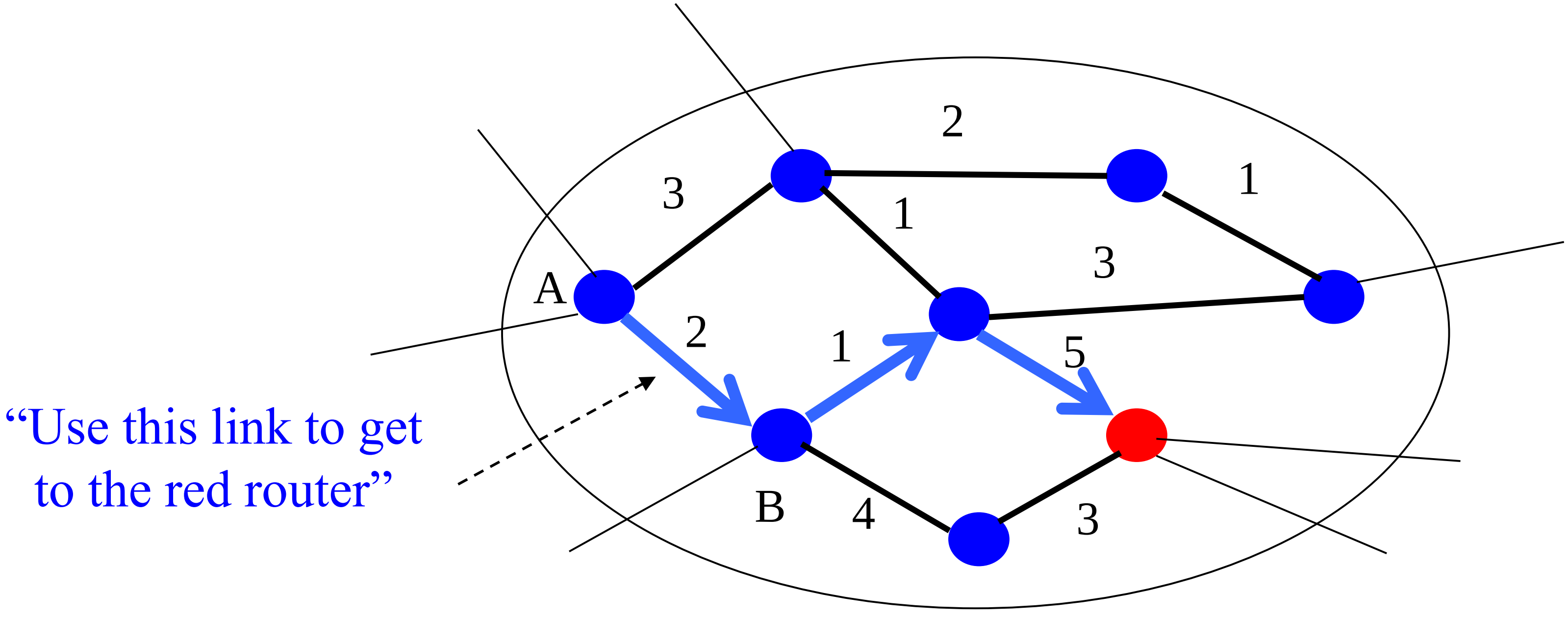
- Interior Gateway Protocols (IGP)
  - RIP, OSPF, IS-IS (similar to OSPF, an ISO standard), ...
  - Only exchange route information within a domain (AS)
  - go out via default gateways
- Exterior Gateway Protocols (EGP)
  - BGP
  - Only exchange route information among domains (AS)

AS, IGP and EGP



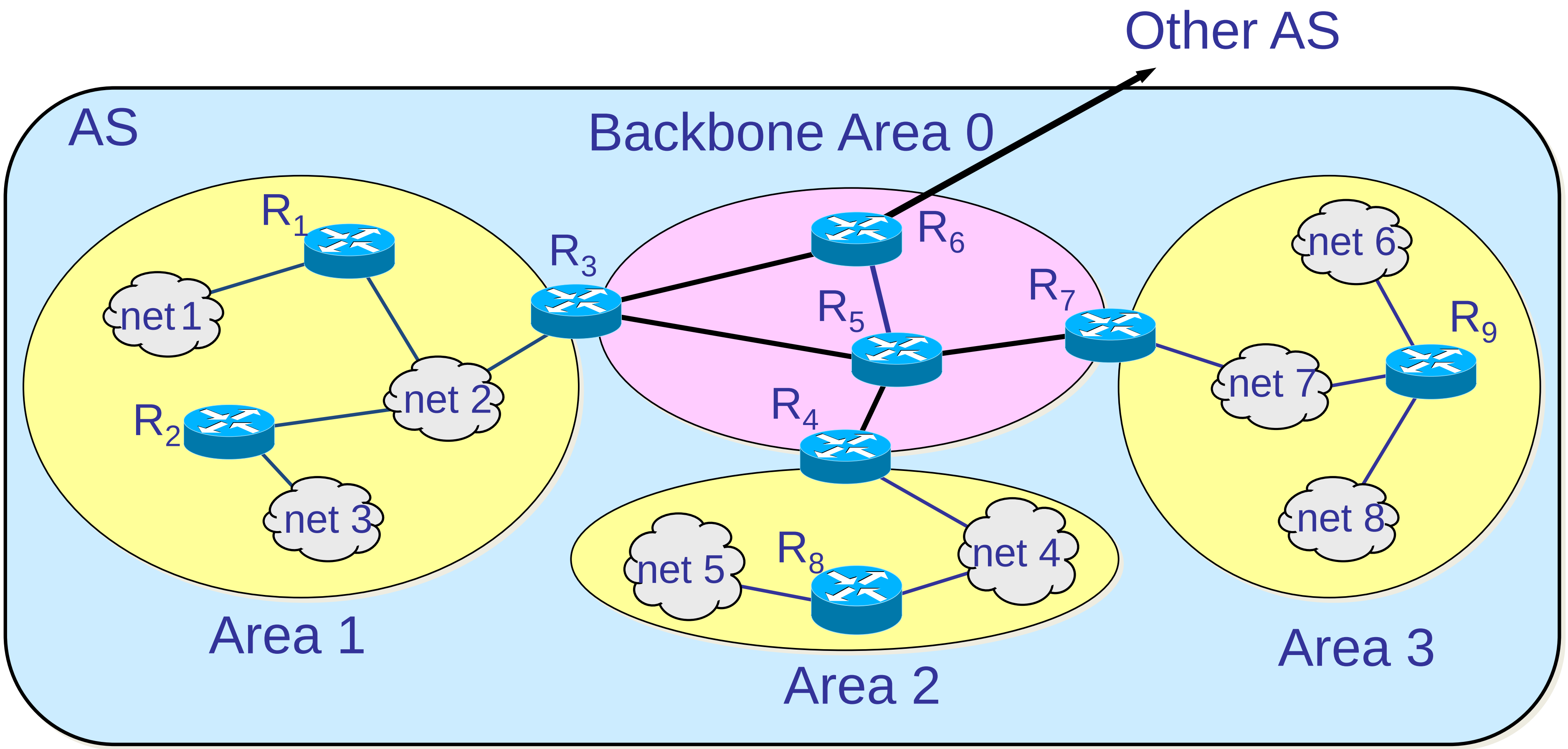
OSPF— Interior Routing Protocol

- OSPF (Open Shortest Path First) is link-state routing:
  - Routers flood information to learn topology
  - Then run Dijkstra to compute routes



OSPF

- OSPF divides one large network (AS) into areas connected to a backbone area
  - Helps to scale; summaries go over area borders

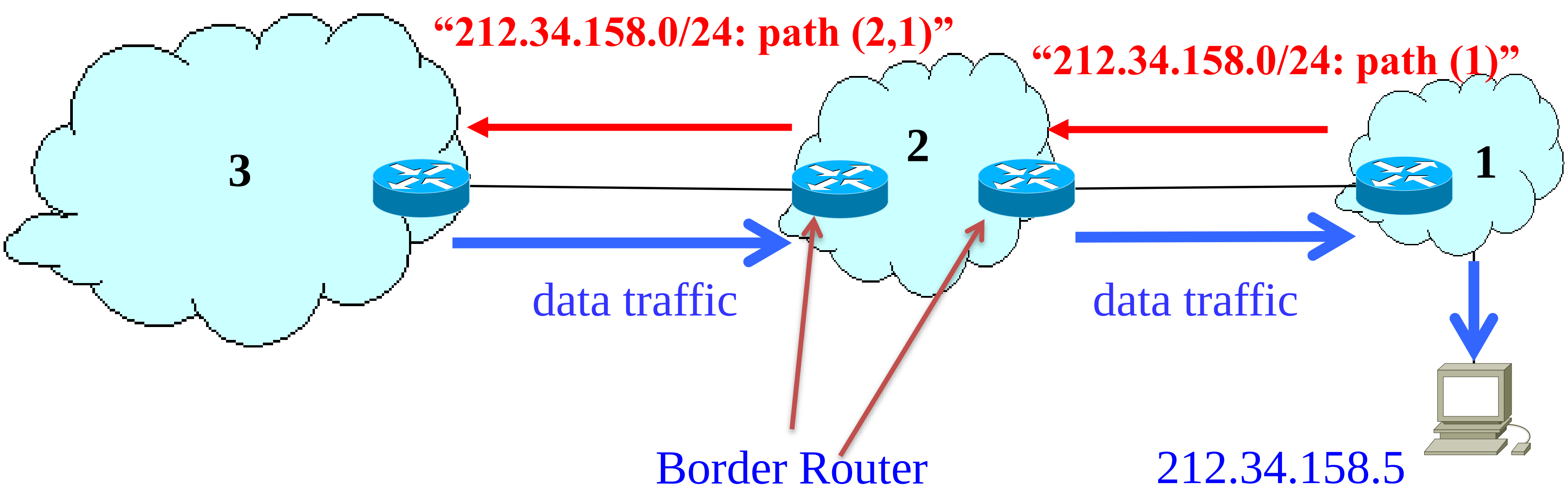


BGP— Exterior Routing Protocol

- BGP (Border Gateway Protocol) computes routes across interconnected ASes
  - Key role is to respect networks' policy constraints
- Example policy constraints:
  - No commercial traffic for educational network
  - Never put Japan on route for traffic of PLA
  - Choose cheaper network
  - Choose better performing network
  - Traffic starting and ending at Tencent shouldn't transit 360

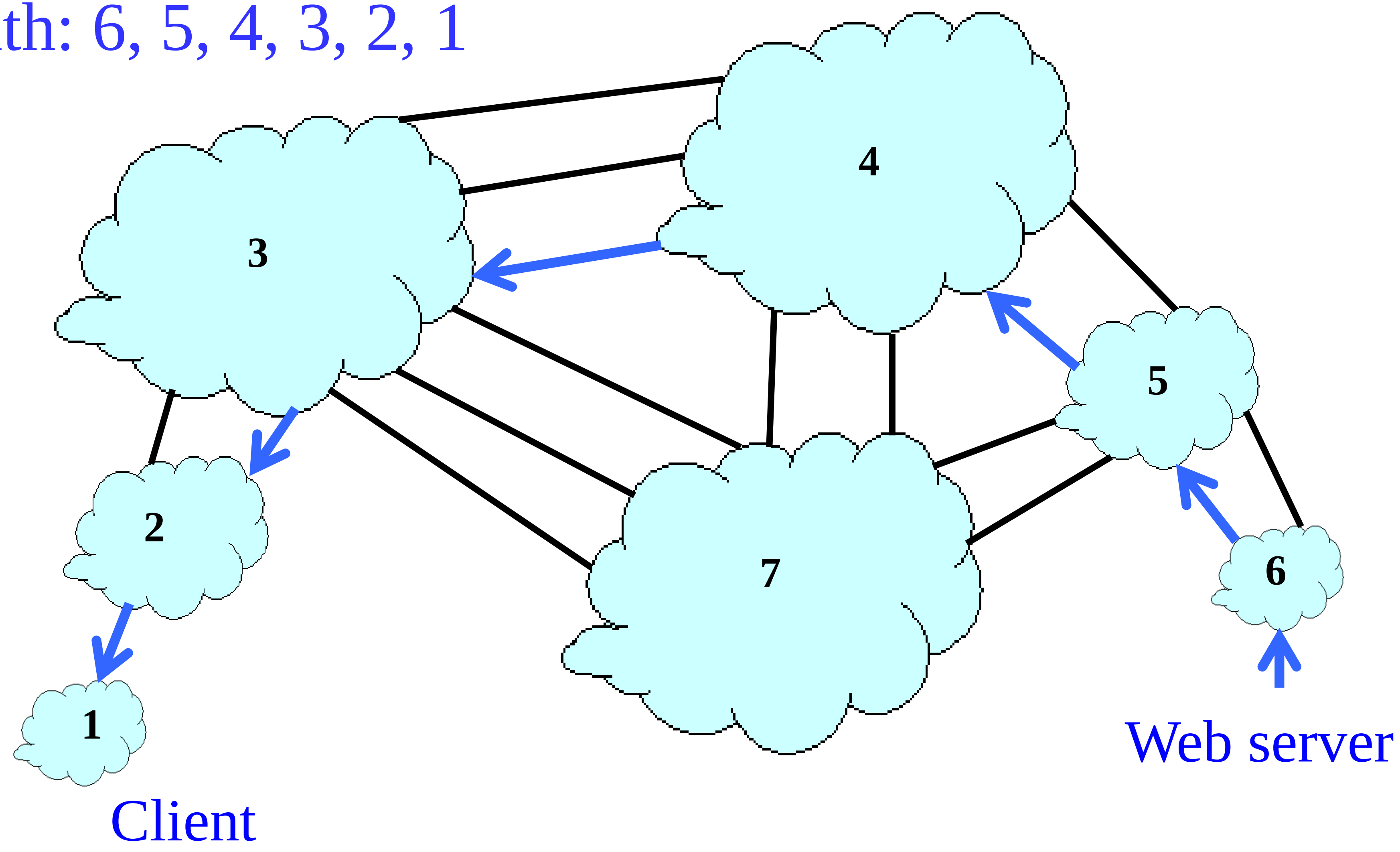
Border Gateway Protocol

- BGP propagates messages along policy-compliant routes
  - ASes exchange info about who they can reach
  - IP prefix: block of destination IP addresses
  - AS path: sequence of ASes along the path



AS Path

Path: 6, 5, 4, 3, 2, 1

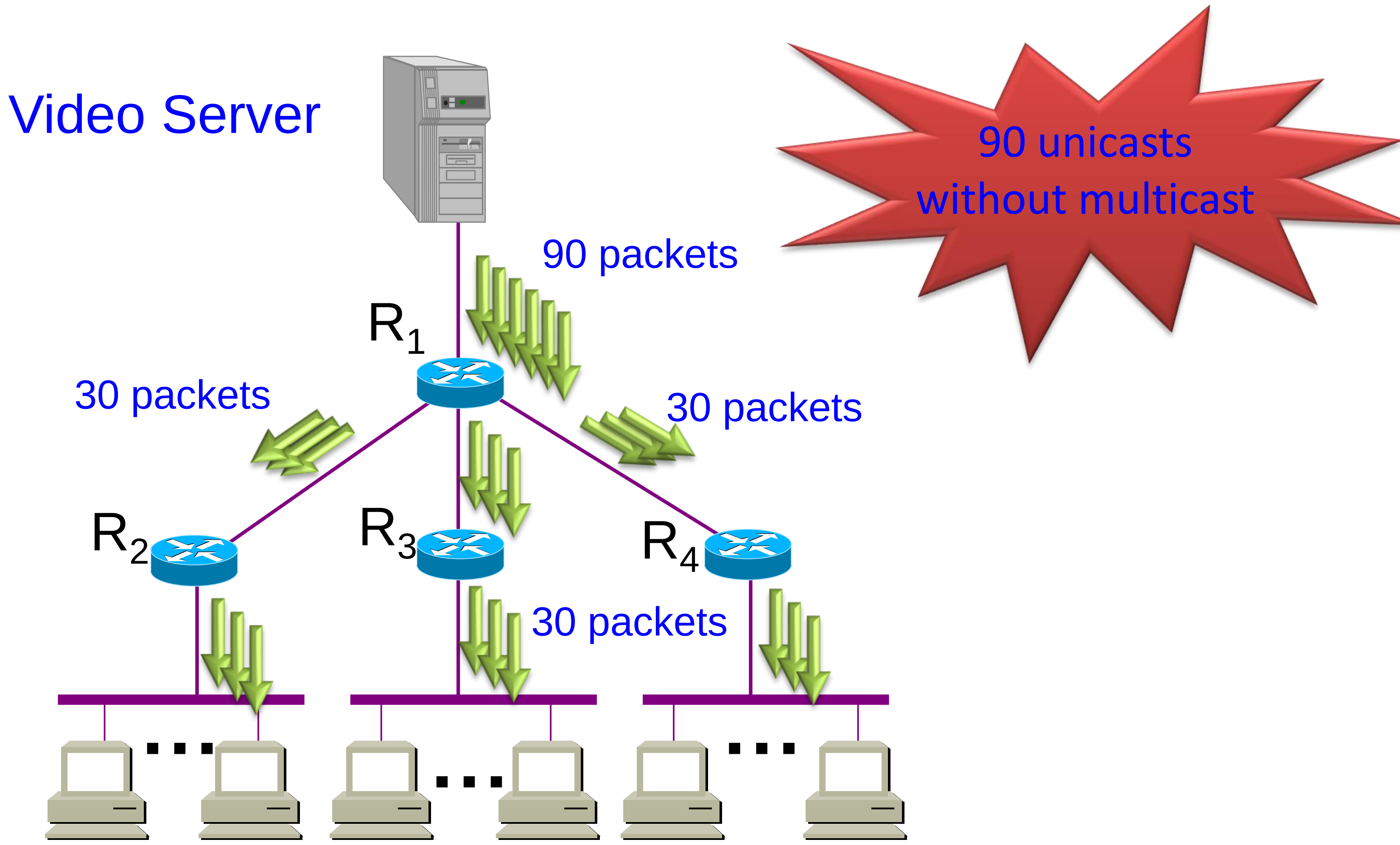


Internet Multicasting

- IP multicasting is not widely used except within a single network, e.g., datacenter, cable TV network.
- Using multicast address (Class D):

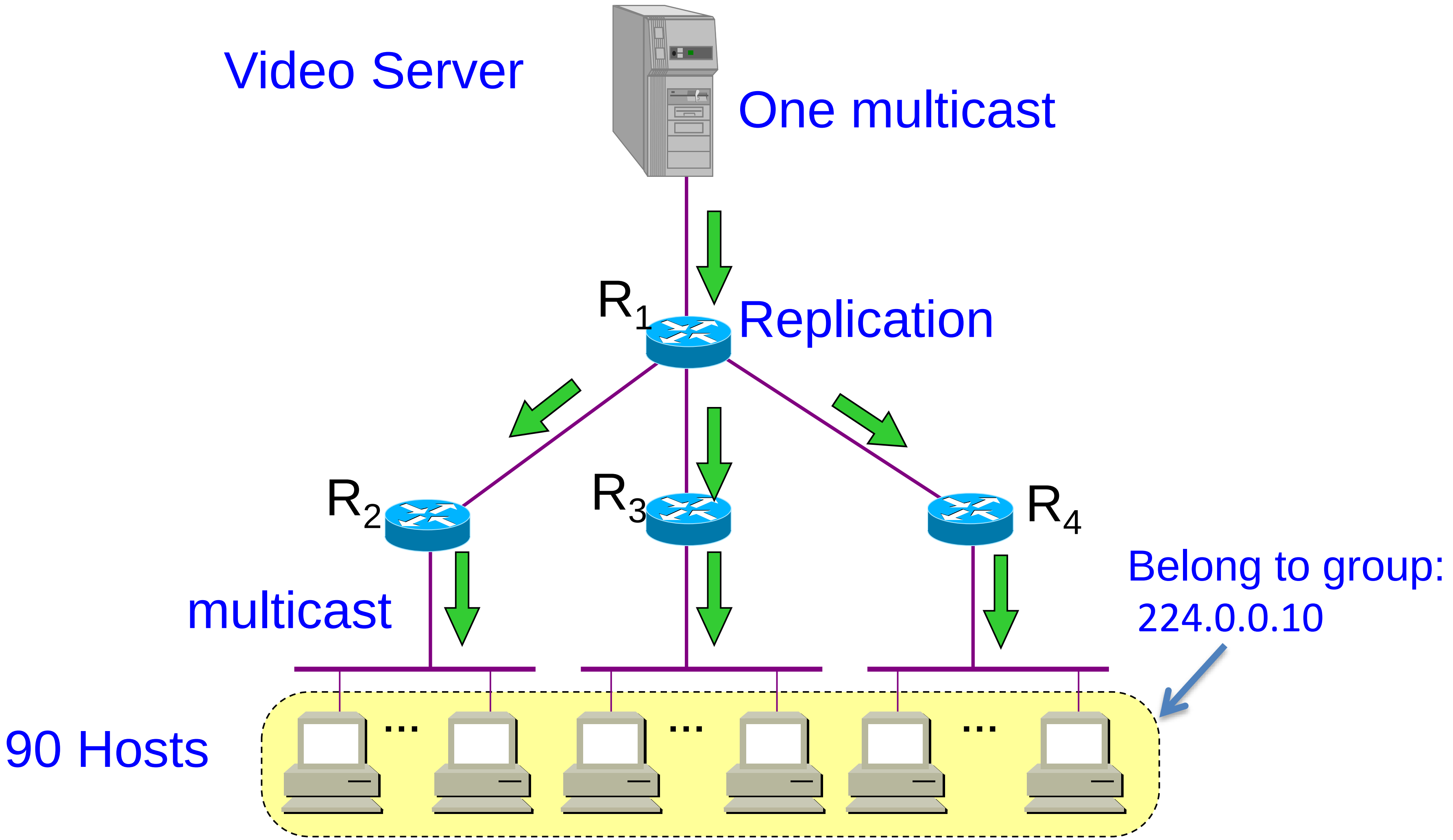
  224.0.0.0 ~ 224.255.255.255

Without IP Multicasting



90 Hosts for the video program

IP Multicasting



Internet Multicasting

- Groups have a reserved IP address range (class D, 224.0.0.0/24)
  - Membership in a group handled by IGMP (Internet Group Management Protocol) that runs at routers
- Routes computed by protocols such as PIM (Protocol Independent Multicast):
  - Dense mode uses RPF (Reverse Path Forwarding) with pruning
  - Sparse mode uses core-based trees

Thank you!

Q & A