

# CS 305 Computer Networks

## Chapter 5 Network Layer – The Control Plane (2)

Jin Zhang

Department of Computer Science and Engineering  
Southern University of Science and Technology

# Chapter 5: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the  
Internet: OSPF

5.4 routing among the ISPs:  
BGP

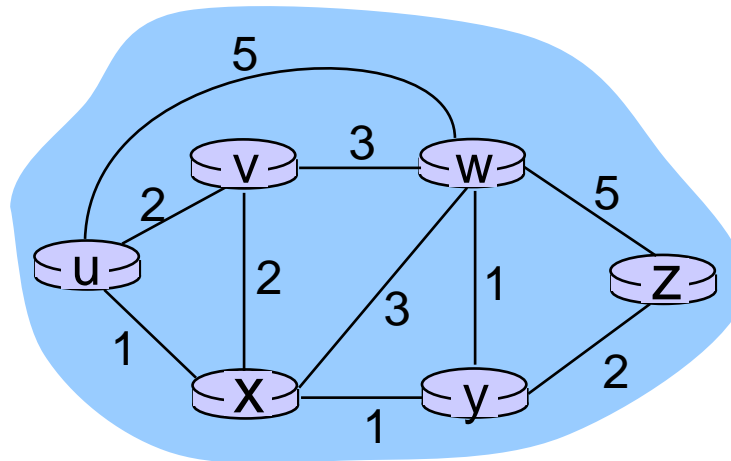
5.5 The SDN control plane

5.6 ICMP: The Internet  
Control Message  
Protocol

5.7 Network management  
and SNMP

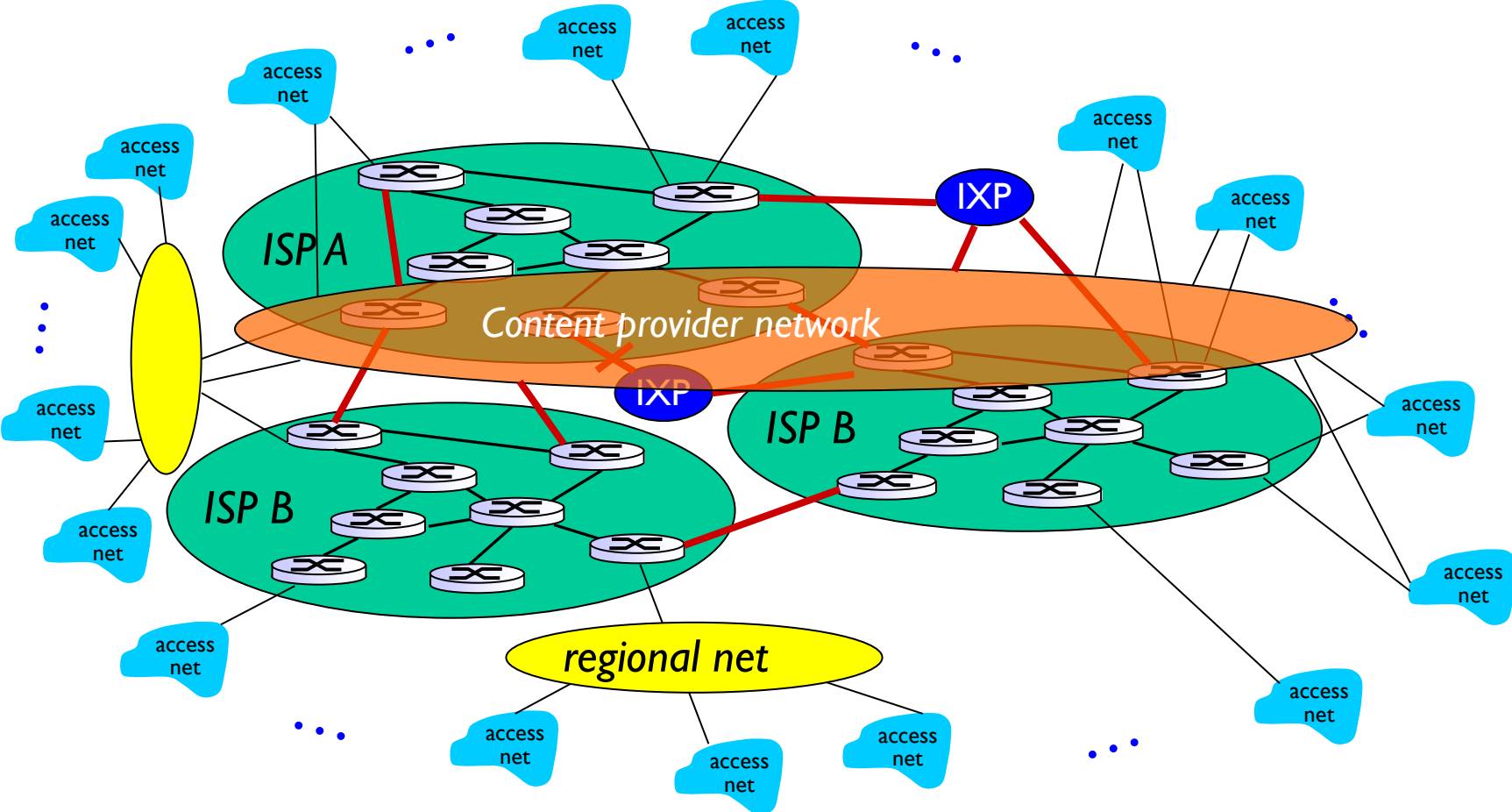
# Review the Link State Routing

Step	N'	D(v),p(v)	D(w),p(w)	D(x),p(x)	D(y),p(y)	D(z),p(z)
0	u	2,u	5,u	1,u	$\infty$	$\infty$
1	ux	2,u	4,x		2,x	$\infty$
2	uxy	2,u	3,y			4,y
3	uxyv		3,y			4,y
4	uxyvw					4,y
5	uxyvwz					



Does the link state routing work on the Internet directly?

---



# The link state routing doesn't work on the Internet!

# Making routing scalable

The link state and distance vector routing studies far is idealized

- all routers identical
- network “flat”

... *not* true in practice

*scale:* with billions of destinations:

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

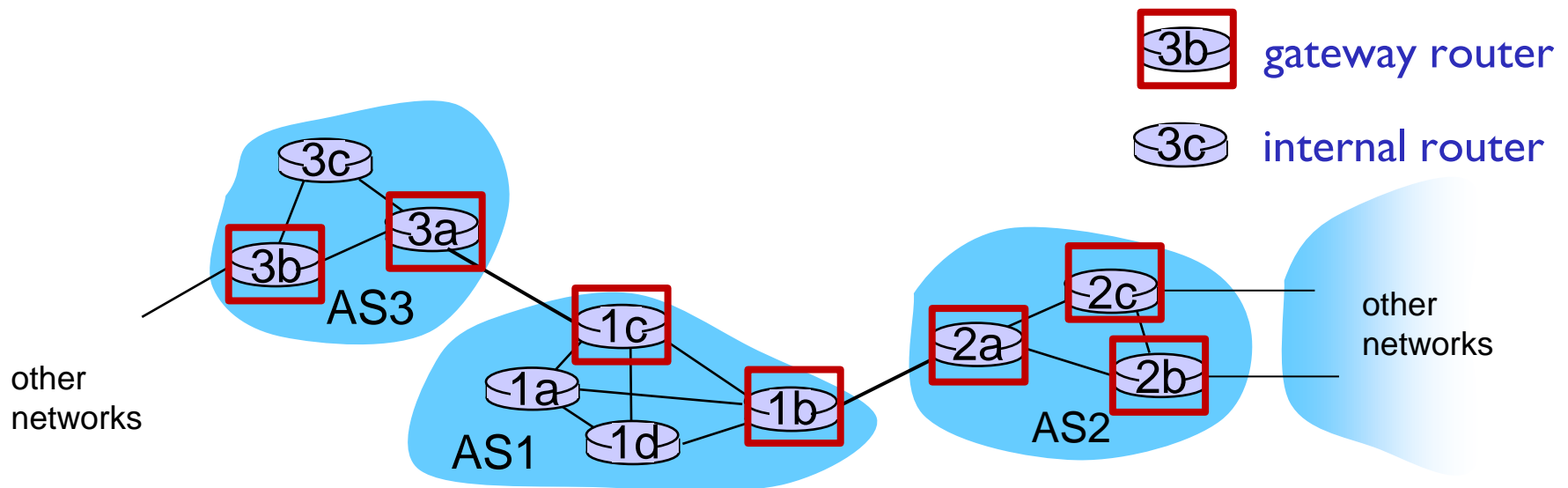
*administrative autonomy*

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

# Internet approach to scalable routing

aggregate routers into regions known as “autonomous systems” (AS) (a.k.a. “domains”)

- Gateway router: at “edge” of its own AS, has link(s) to router(s) in other AS
- Interior router: no link to other AS



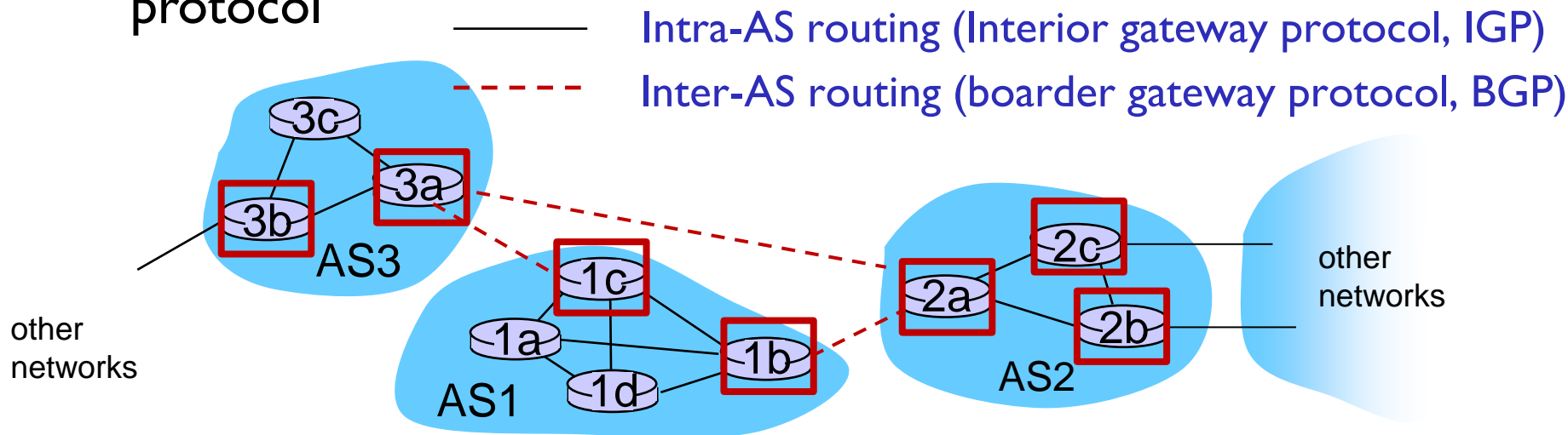
# Internet approach to scalable routing

## intra-AS routing

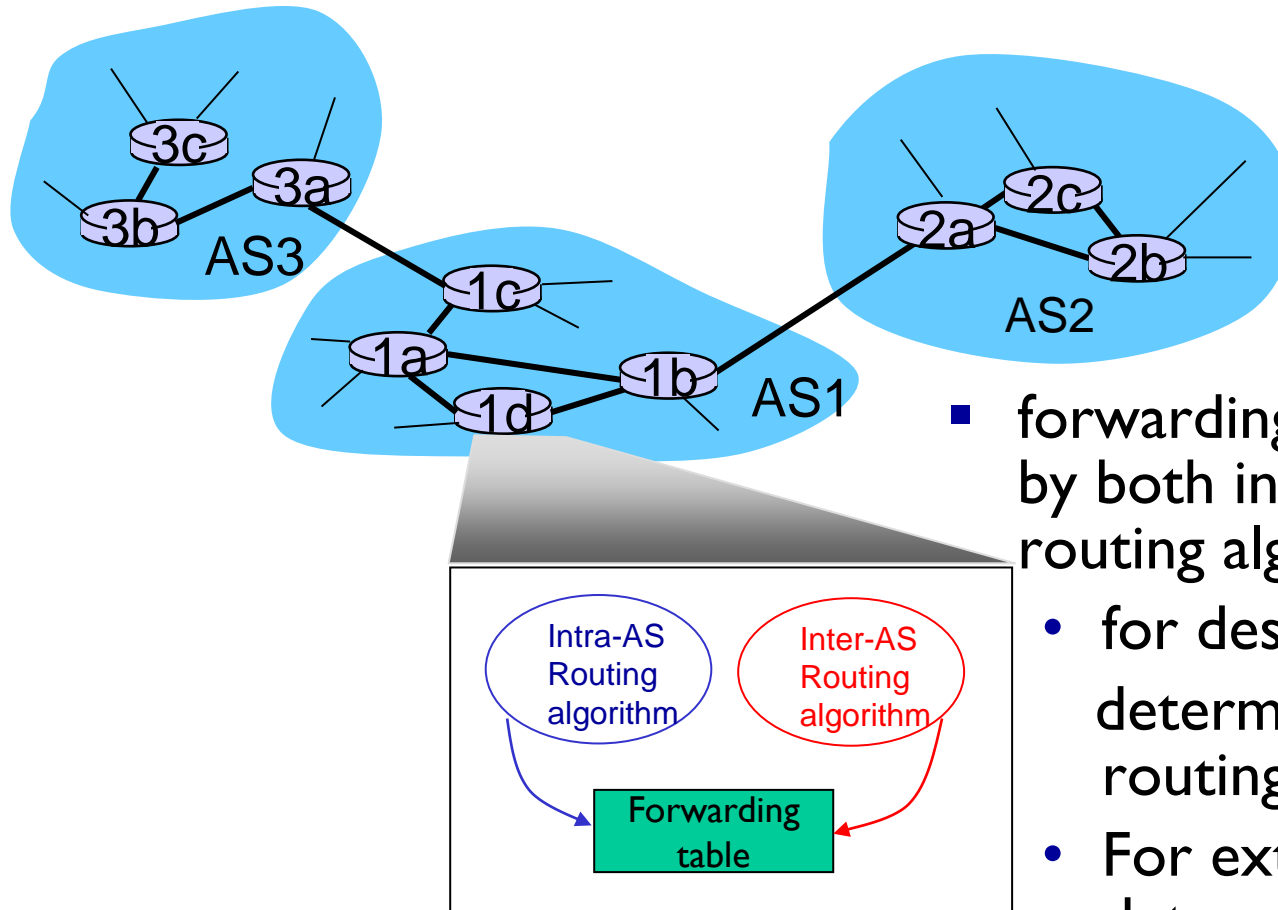
- routing among hosts, routers in same AS (“network”)
- all routers in AS must run *same* intra-domain protocol
- routers in *different* AS can run *different* intra-domain routing protocol

## inter-AS routing

- routing among AS'es
- gateways perform inter-domain routing (as well as intra-domain routing)



# Interconnected ASes



- forwarding table configured by both intra- and inter-AS routing algorithm
  - for destinations within AS: determined by intra-AS routing
  - For external destinations: determined by both inter-AS & intra-AS routing



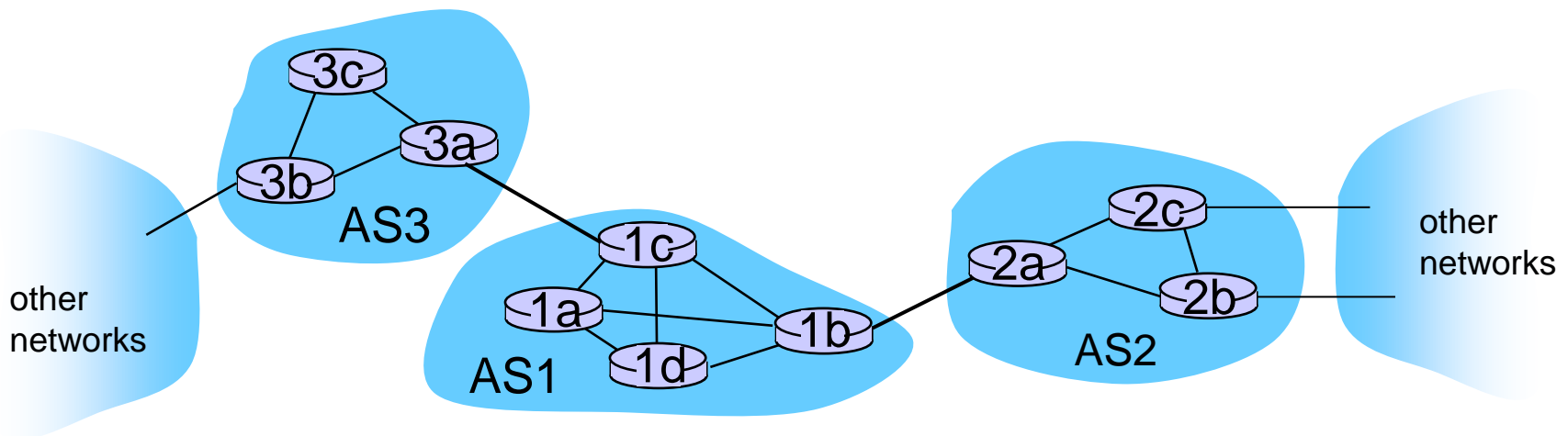
# Inter-AS tasks

- suppose router in AS1 receives datagram destined outside of AS1:
  - router should forward packet to gateway router, but which one?

*AS1 must:*

1. learn which destds are reachable through AS2, which through AS3
2. propagate this reachability info to all routers in AS1

*job of inter-AS routing!*



# Intra-AS Routing

- also known as *interior gateway protocols (IGP)*
- most common intra-AS routing protocols:
  - RIP: Routing Information Protocol (distance vector-based)
  - OSPF: Open Shortest Path First (link state-based)
  - IS-IS protocol essentially same as OSPF
  - IGRP: Interior Gateway Routing Protocol (Cisco proprietary for decades, until 2016)

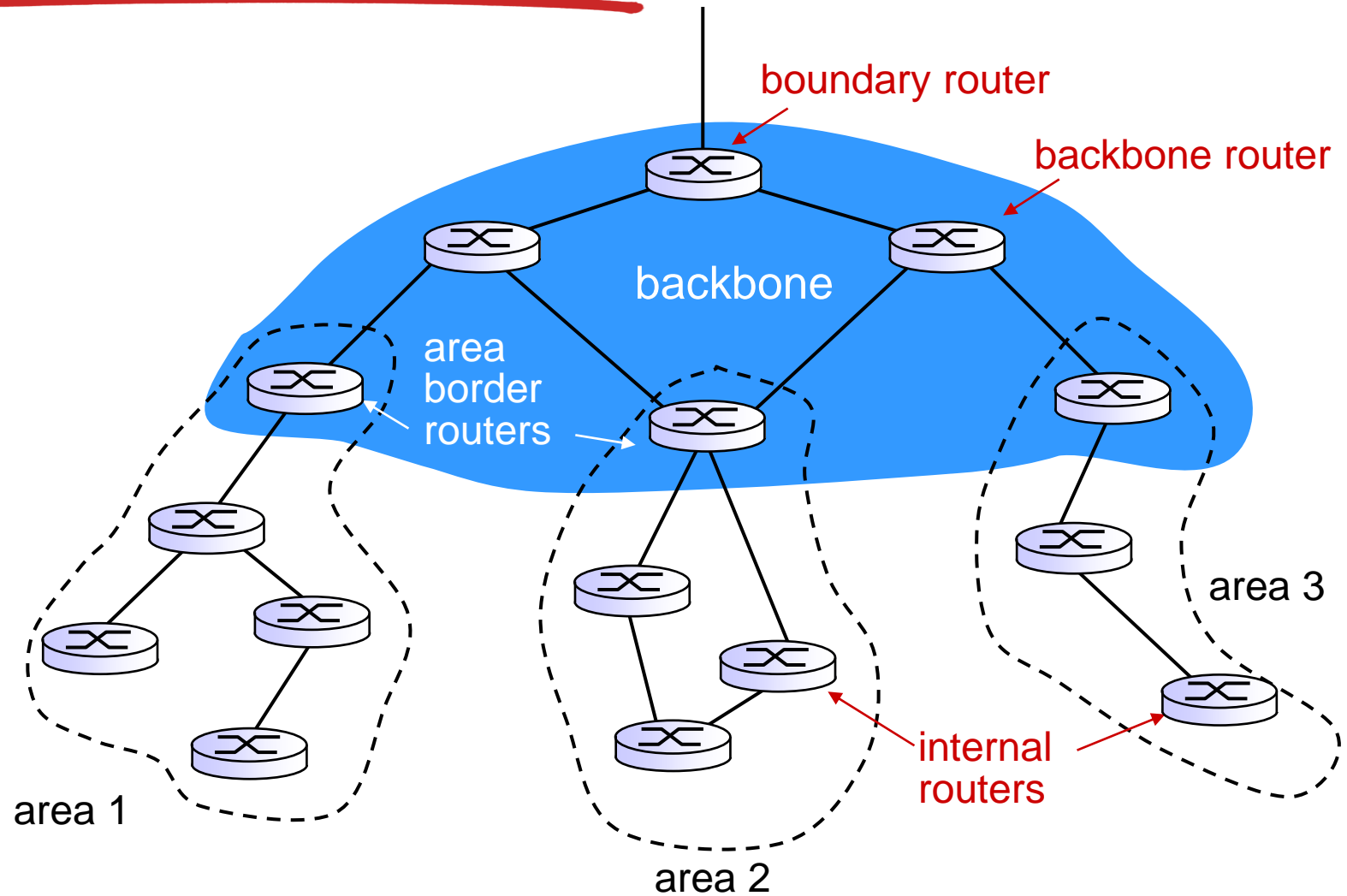
# OSPF (Open Shortest Path First)

- “open”: publicly available
- uses link-state algorithm
  - link state packet dissemination
  - topology map at each node
  - route computation using Dijkstra’s algorithm
- router floods OSPF link-state advertisements to all other routers in *entire* AS
  - carried in OSPF messages directly over IP (rather than TCP or UDP)
  - link state: for each attached link

# OSPF “advanced” features

- **security**: all OSPF messages authenticated (to prevent malicious intrusion)
- **multiple** same-cost **paths** allowed (only one path in RIP)
- integrated uni- and **multi-cast** support:
  - Multicast OSPF (MOSPF) uses same topology data base as OSPF
- **hierarchical** OSPF in large domains.

# Hierarchical OSPF



# Hierarchical OSPF

- *two-level hierarchy*: local area, backbone.
  - link-state advertisements only in area
  - each nodes has detailed area topology; only know direction (shortest path) to nets in other areas.
- *area border routers*: “summarize” distances to nets in own area, advertise to other Area Border routers.
- *backbone routers*: run OSPF routing limited to backbone.
- *boundary routers*: connect to other AS' es.

# Chapter 5: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the Internet: OSPF

5.4 routing among the ISPs:  
BGP

5.5 The SDN control plane

5.6 ICMP: The Internet  
Control Message  
Protocol

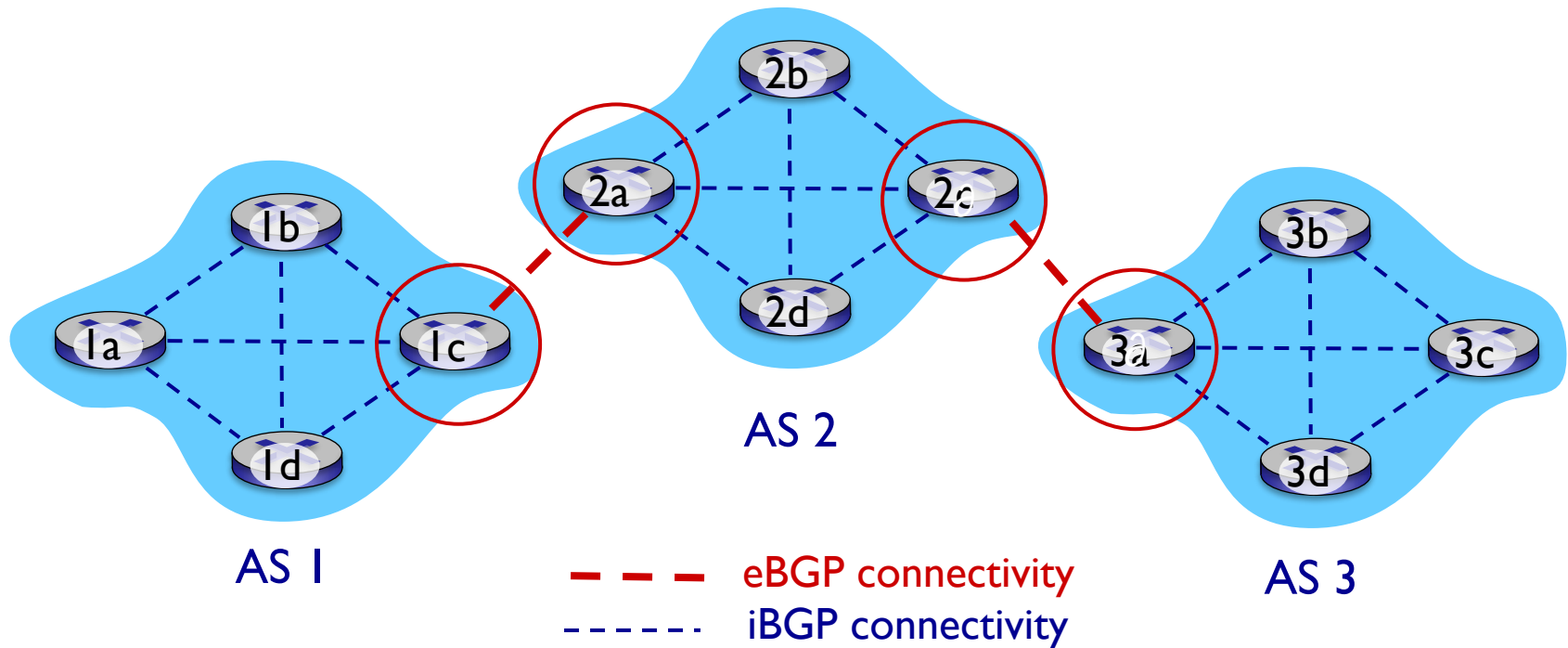
5.7 Network management  
and SNMP

# Internet inter-AS routing: BGP

- **BGP (Border Gateway Protocol):** *the de facto inter-domain routing protocol*
  - “glue that holds the Internet together”
- Main functions BGP provides :
  - obtain subnet reachability information from neighboring Ases: **eBGP**
  - propagate reachability information to all AS-internal routers: **iBGP**
  - determine “good” routes to other networks based on reachability information and *policy*
- allows subnet to advertise its existence to rest of Internet: *“I am here”*



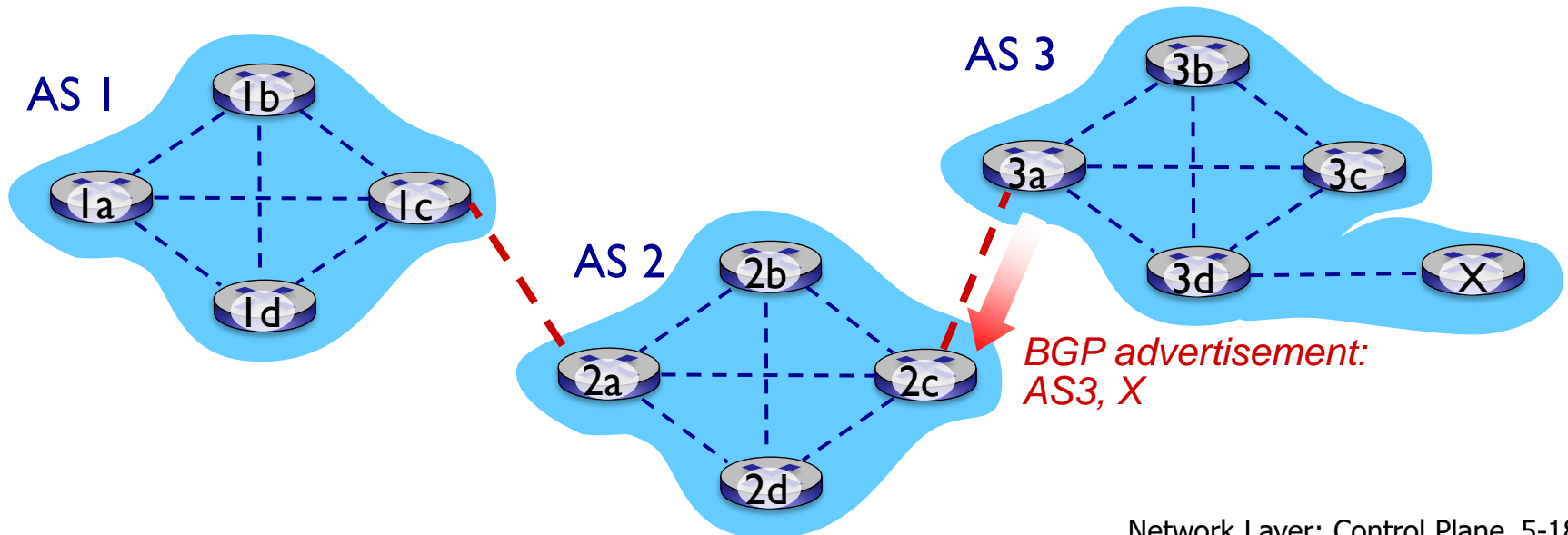
# eBGP, iBGP connections



gateway routers run both eBGP and iBGP protocols

# BGP basics

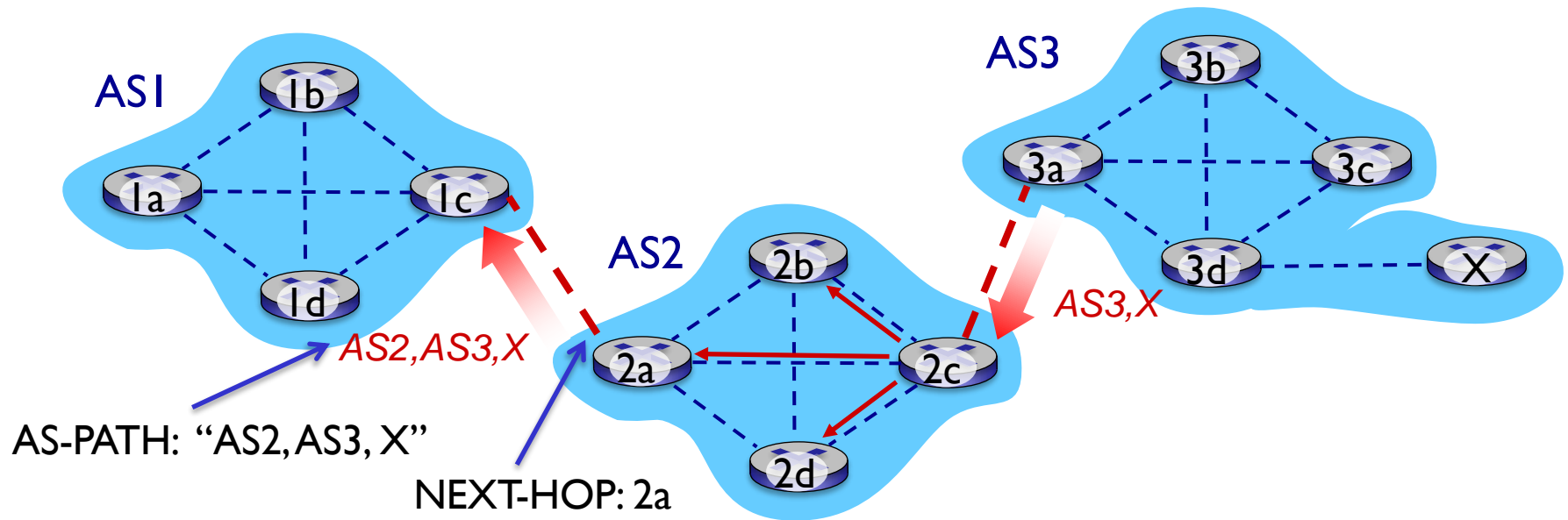
- **BGP session:** two BGP routers (“peers”) exchange BGP messages over semi-permanent TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a “path vector” protocol)
- when AS3 gateway router 3a advertises path **AS3,X** to AS2 gateway router 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X



# Path attributes and BGP routes

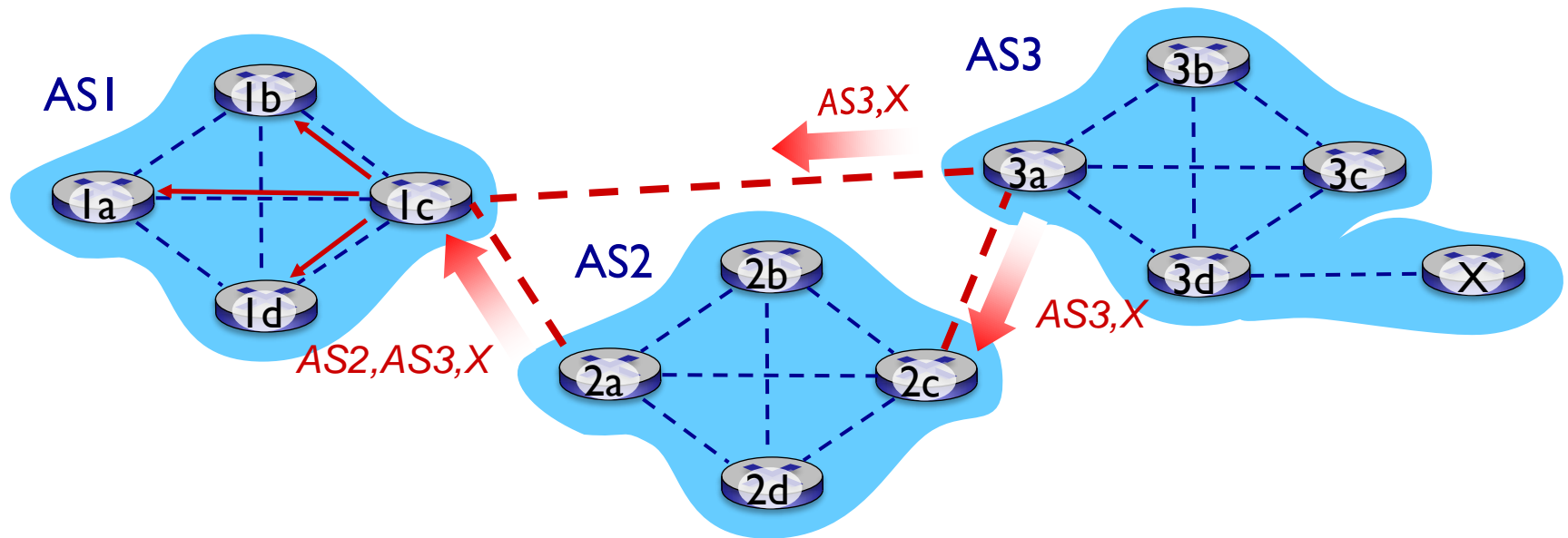
- advertised prefix includes BGP attributes
  - Prefix (destination) + attributes = “route”
- two important attributes:
  - **AS-PATH**: list of ASes through which the advertisement has passed
  - **NEXT-HOP**: IP address of the router interface that begins the AS-PATH
- *Policy-based routing*:
  - gateway receiving route advertisement uses *import policy* to accept/decline path (e.g., never route through AS Y).
  - AS policy also determines whether to *advertise* path to other neighboring ASes

# BGP path advertisement



- AS2 router 2c receives path advertisement **AS3,X** (via eBGP) from AS3 router 3a
- Based on AS2 policy, AS2 router 2c accepts path AS3,X, propagates (via iBGP) to all AS2 routers
- Based on AS2 policy, AS2 router 2a advertises (via eBGP) path **AS2, AS3,X** to AS1 router 1c

# BGP path advertisement

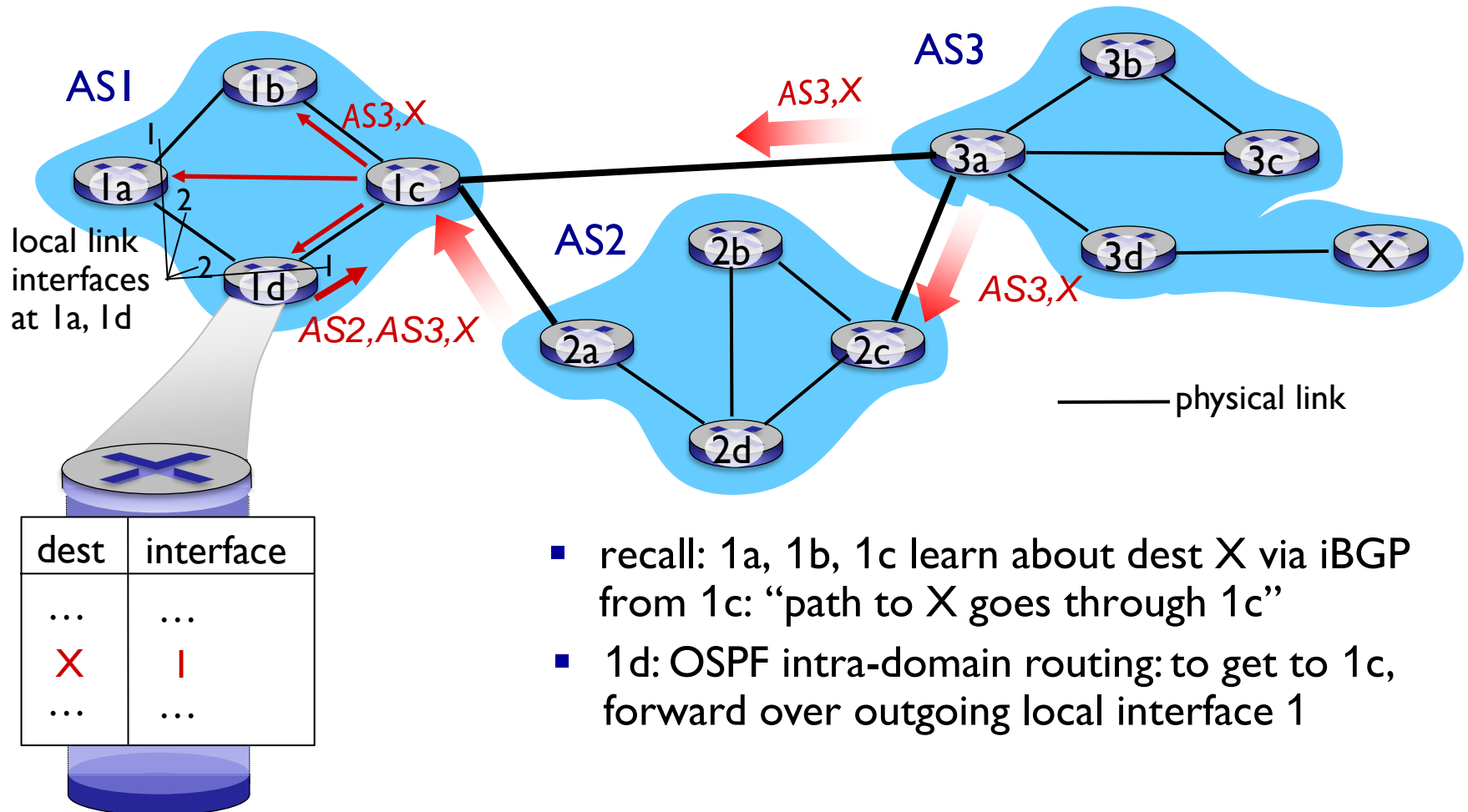


gateway router may learn about **multiple** paths to destination:

- AS1 gateway router 1c learns path **AS2,AS3,X** from 2a
- AS1 gateway router 1c learns path **AS3,X** from 3a
- Based on policy (shortest AS PATH), AS1 gateway router 1c chooses path **AS3,X**, and advertises *path within AS 1 via iBGP*

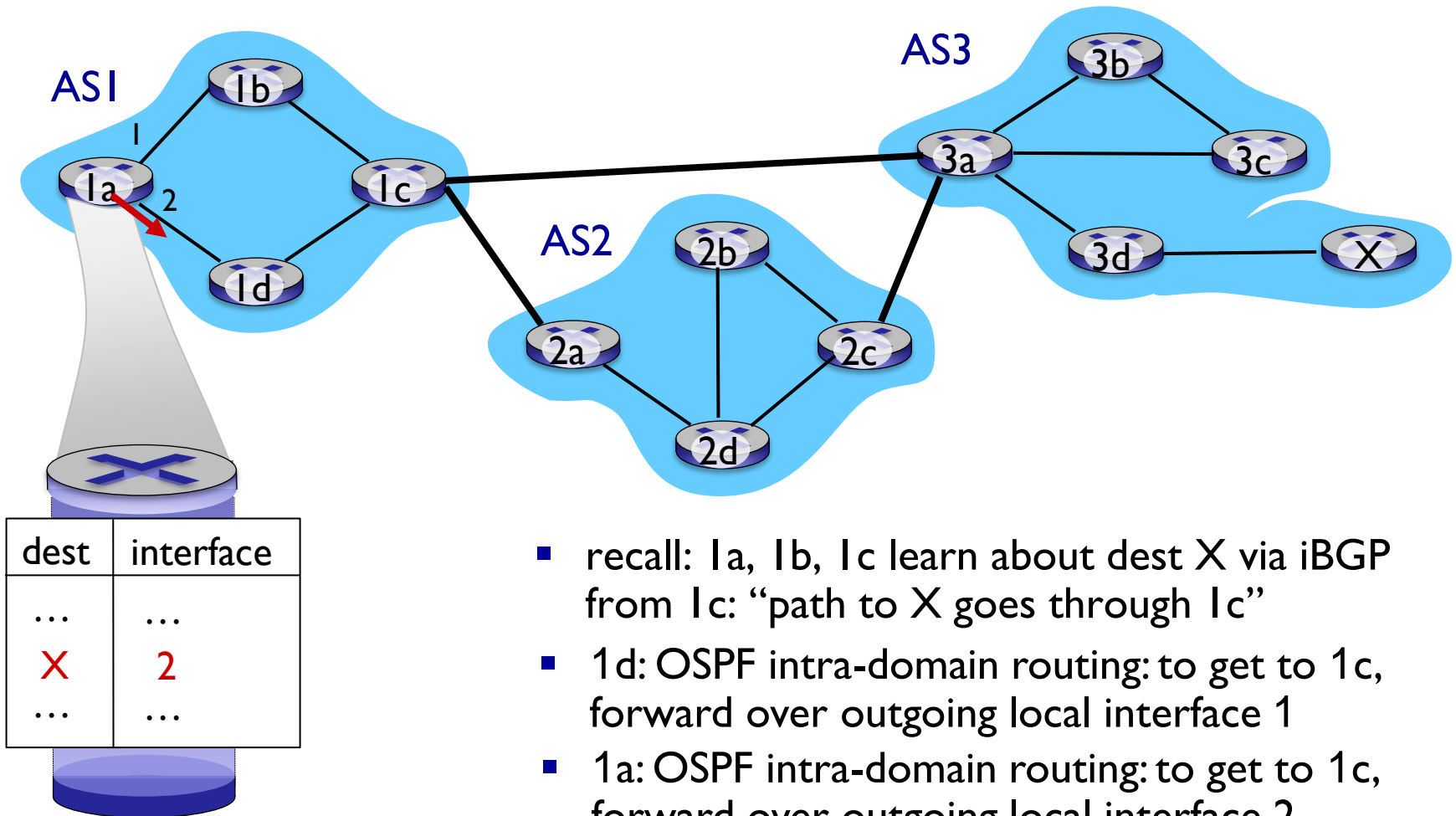
# BGP, OSPF, forwarding table entries

Q: how does router set forwarding table entry to distant prefix?



# BGP, OSPF, forwarding table entries

Q: how does router set forwarding table entry to distant prefix?



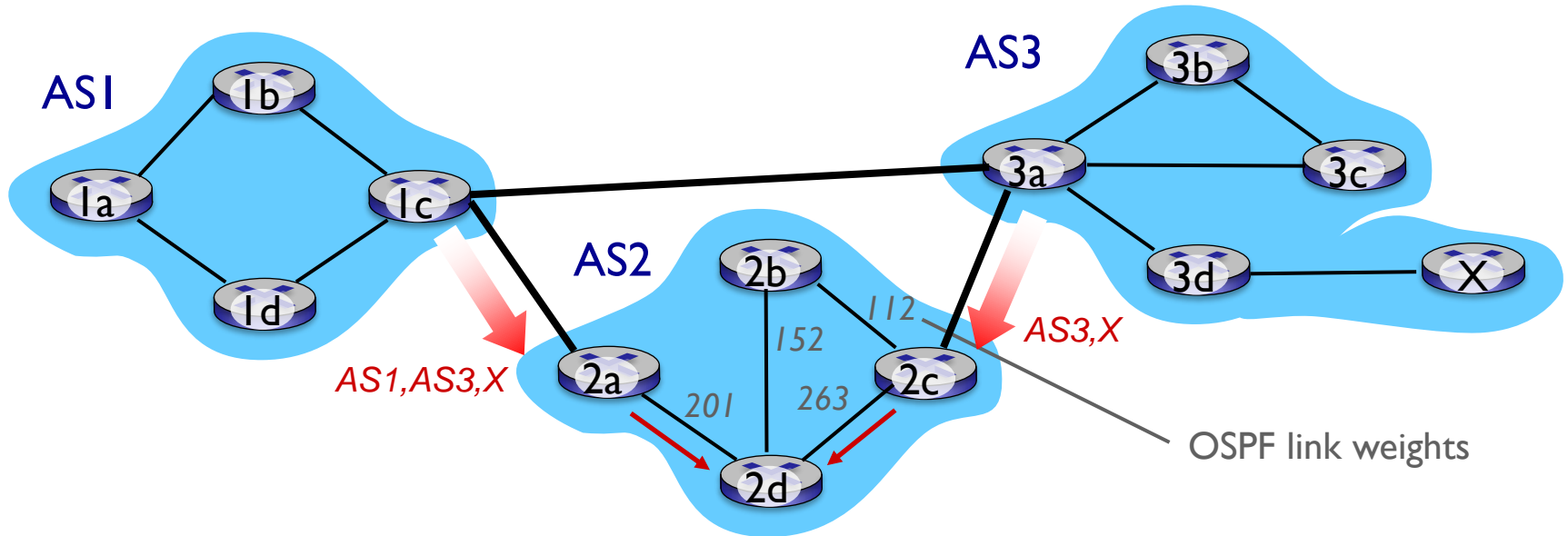
- recall: 1a, 1b, 1c learn about dest X via iBGP from 1c: “path to X goes through 1c”
- 1d: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 1
- 1a: OSPF intra-domain routing: to get to 1c, forward over outgoing local interface 2

# BGP route selection

- router may learn about more than one route to destination AS, selects route based on:
  1. local preference value attribute: policy decision
  2. shortest AS-PATH
  3. closest NEXT-HOP router: hot potato routing
  4. additional criteria

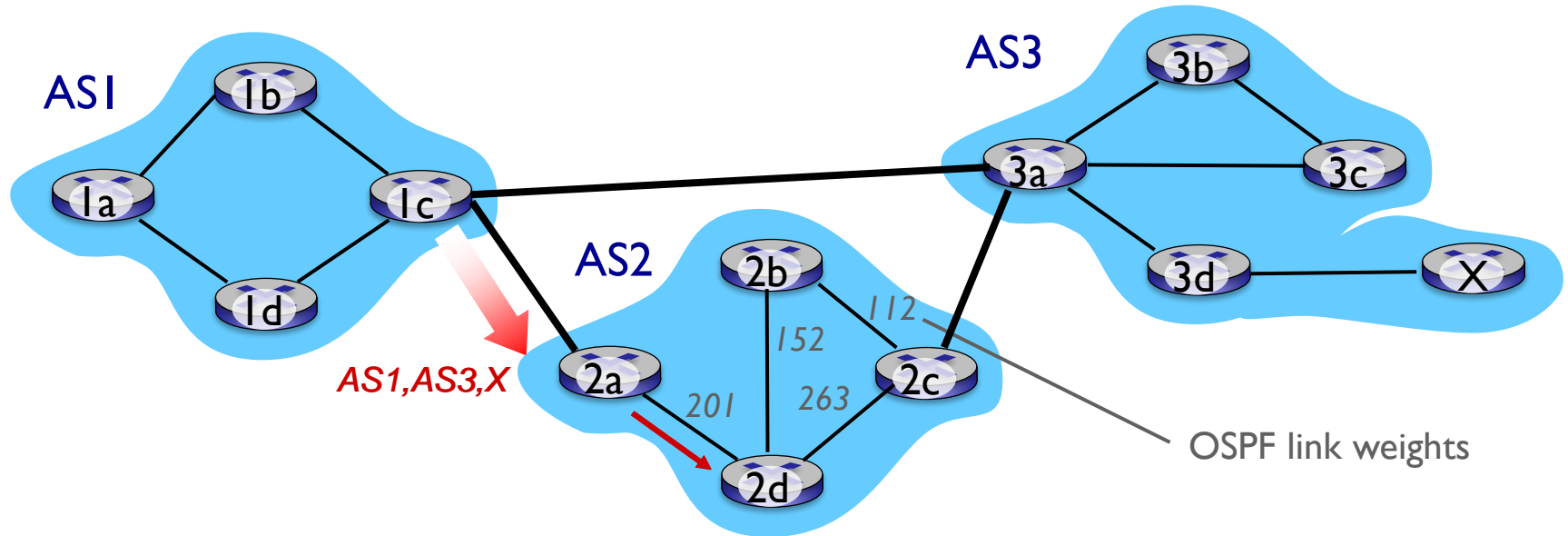


# Hot Potato Routing



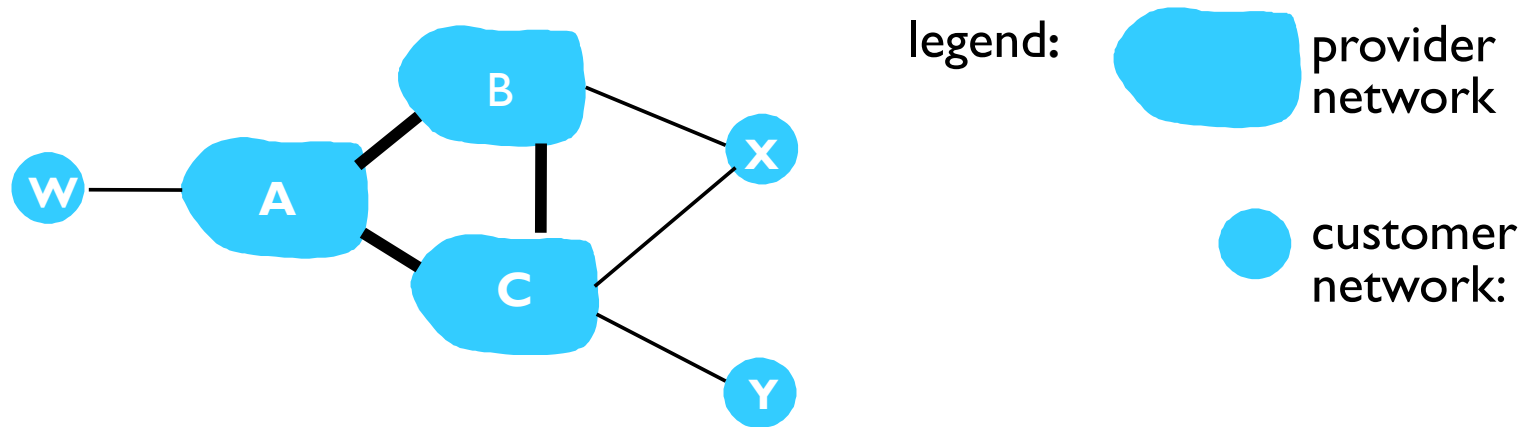
- 2d learns (via iBGP) it can route to X via 2a or 2c

# Hot Potato Routing



- 2d learns (via iBGP) it can route to X via 2a or 2c
- *hot potato routing*: choose local gateway that has least intra-domain cost (e.g., 2d chooses 2a, even though more AS hops to X): don't worry about inter-domain cost!

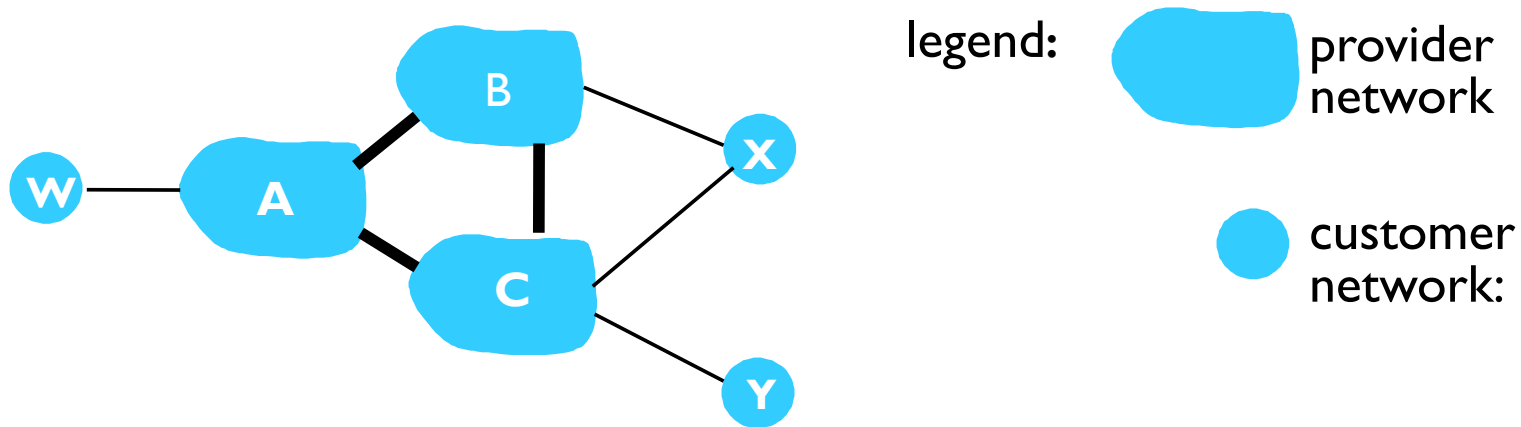
# BGP: achieving policy via advertisements



Suppose an ISP only wants to route traffic to/from its customer networks  
(does not want to carry transit traffic between other ISPs)

- A advertises path A<sub>w</sub> to B and to C
- B *chooses not to advertise* B<sub>A<sub>w</sub></sub> to C:
  - B gets no “revenue” for routing C<sub>B<sub>A<sub>w</sub></sub></sub>, since none of C, A, w are B’s customers
  - C does not learn about C<sub>B<sub>A<sub>w</sub></sub></sub> path
- C will route C<sub>A<sub>w</sub></sub> (not using B) to get to w

# BGP: achieving policy via advertisements



Suppose an ISP only wants to route traffic to/from its customer networks  
(does not want to carry transit traffic between other ISPs)

- A,B,C are *provider networks*
- X,W,Y are customer (of provider networks)
- X is **dual-homed**: attached to two networks
- **policy to enforce**: X does not want to route from B to C via X
  - .. so X will not advertise to B a route to C

# Why different Intra-, Inter-AS routing ?

## *policy:*

- inter-AS: admin wants control over how its traffic routed, who routes through its net.
- intra-AS: single admin, so no policy decisions needed

## *scale:*

- hierarchical routing saves table size, reduced update traffic

## *performance:*

- intra-AS: can focus on performance
- inter-AS: policy may dominate over performance

# Chapter 5: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the Internet: OSPF

5.4 routing among the ISPs: BGP

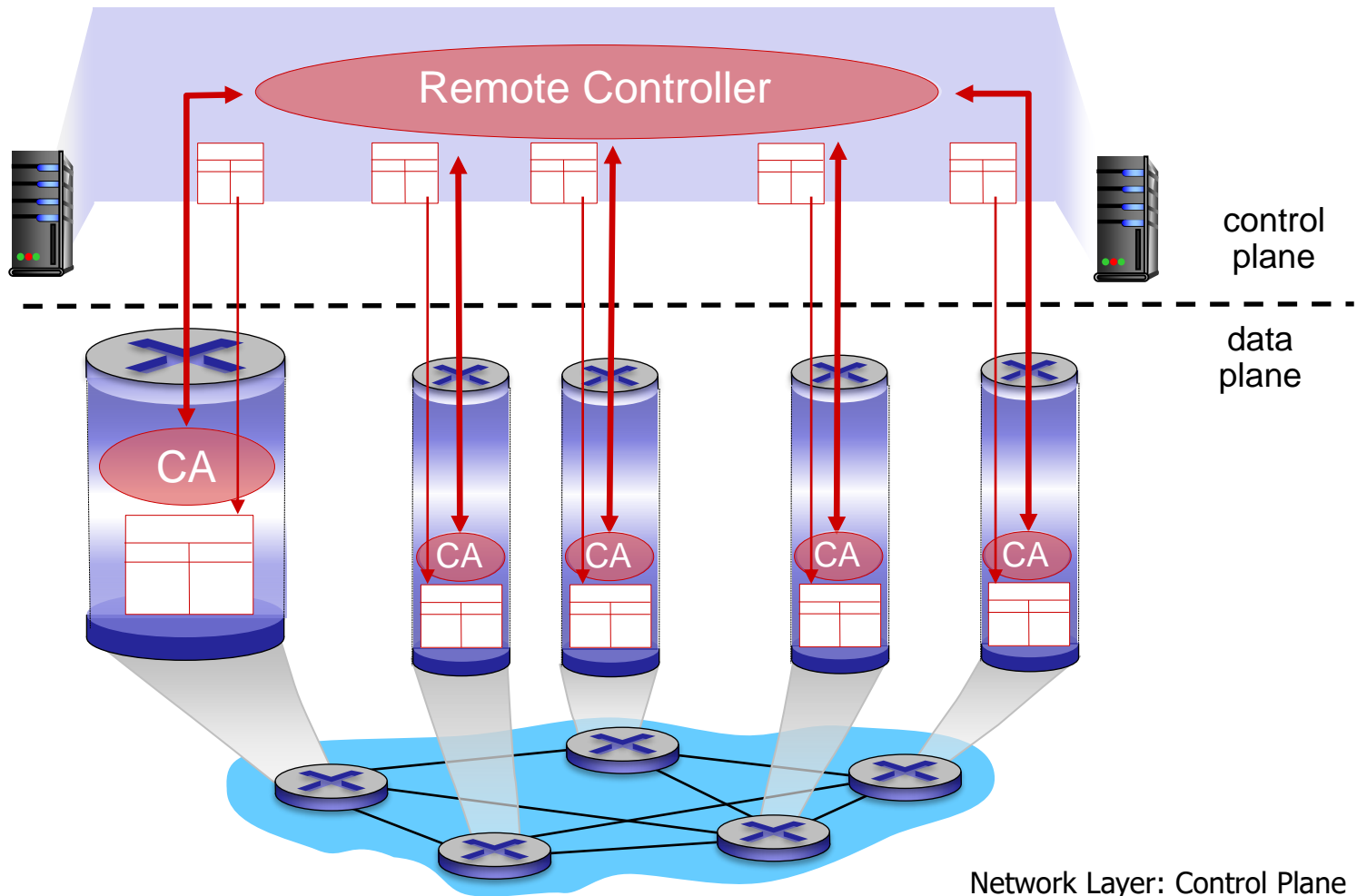
5.5 The SDN control plane

5.6 ICMP: The Internet Control Message Protocol

5.7 Network management and SNMP

# Recall: SDN logically centralized control plane

A distinct (typically remote) controller interacts with local control agents (CAs) in routers to compute forwarding tables



# Software defined networking (SDN)

*Why* a *logically centralized* control plane?

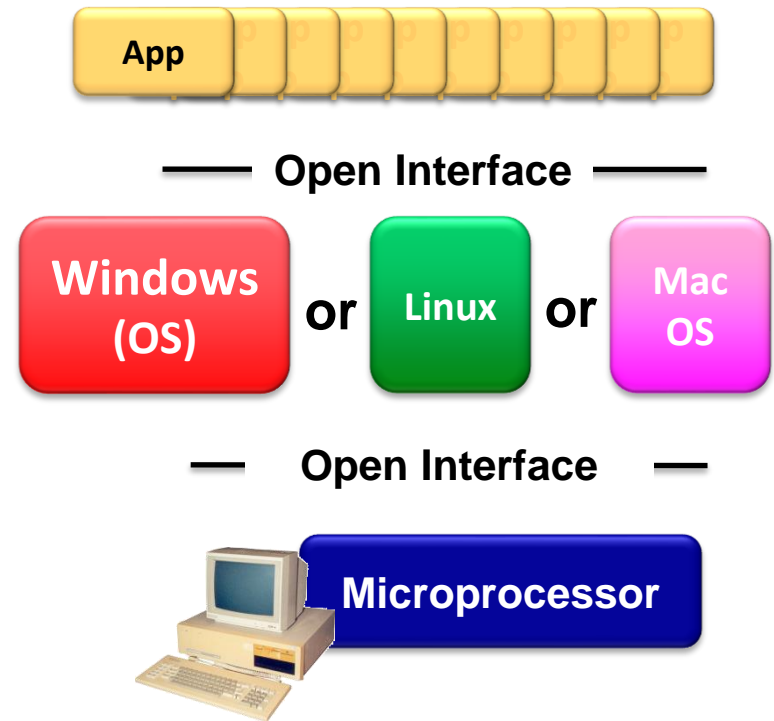
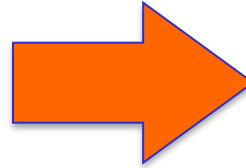
- easier network management: avoid router misconfigurations, greater flexibility of traffic flows
- table-based forwarding (recall OpenFlow API) allows “programming” routers
  - centralized “programming” easier: compute tables centrally and distribute
  - distributed “programming” more difficult: compute tables as result of distributed algorithm (protocol) implemented in each and every router
- open (non-proprietary) implementation of control plane



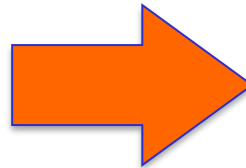
# Analogy: mainframe to PC evolution\*



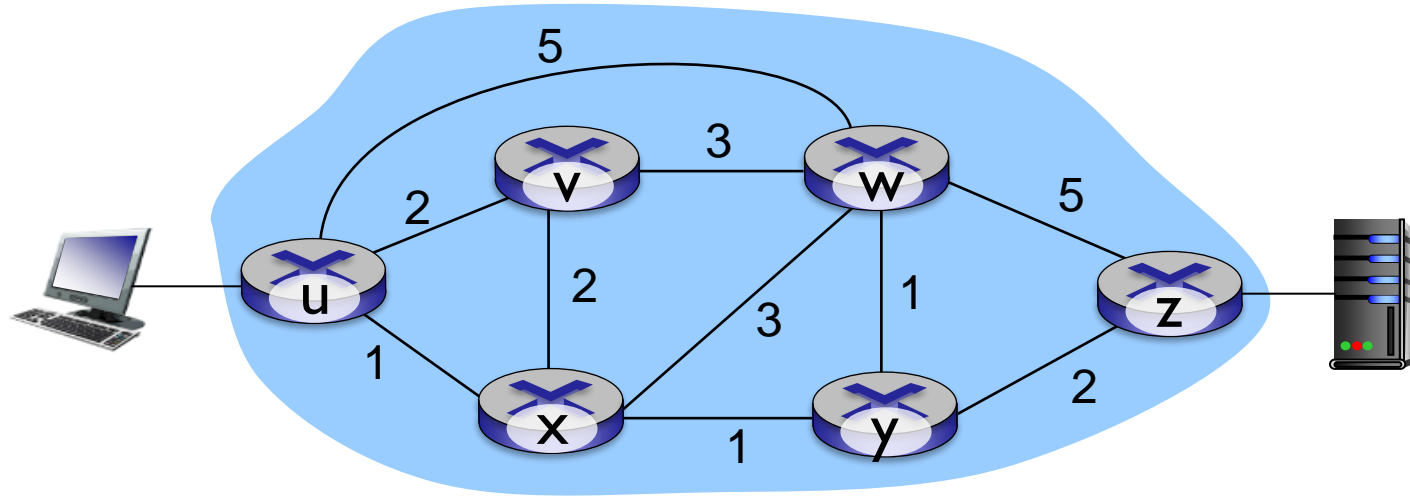
Vertically integrated  
Closed, proprietary  
Slow innovation  
Small industry



Horizontal  
Open interfaces  
Rapid innovation  
Huge industry



# Traffic engineering: difficult traditional routing

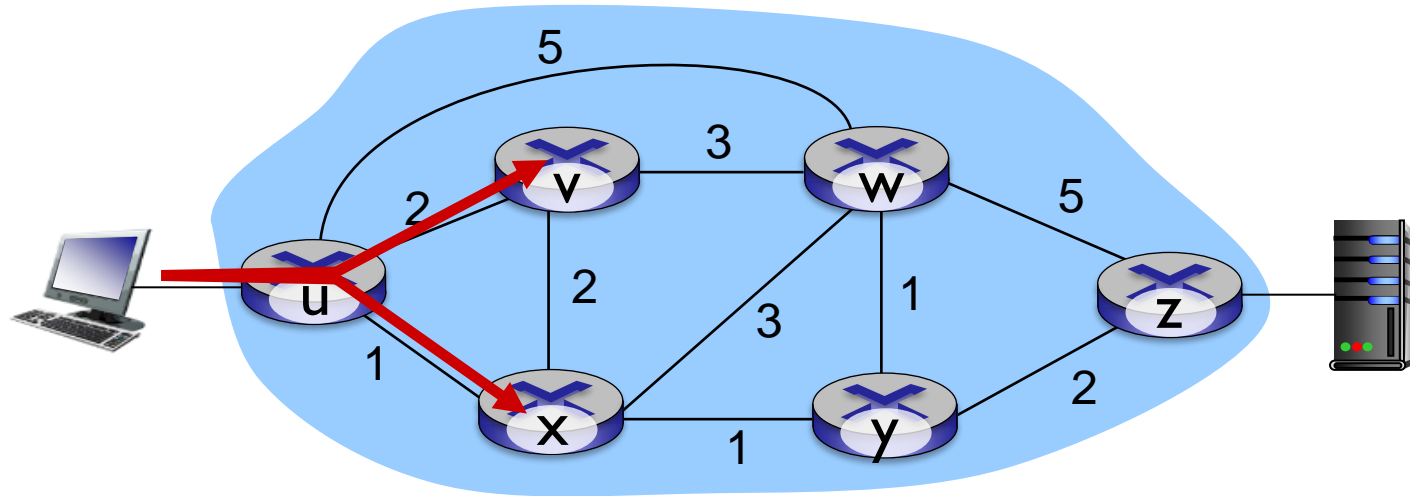


Q: what if network operator wants u-to-z traffic to flow along *uvwz*, x-to-z traffic to flow *xwyz*?

A: need to define link weights so traffic routing algorithm computes routes accordingly (or need a new routing algorithm)!

But the link weights cannot be directly set to certain number

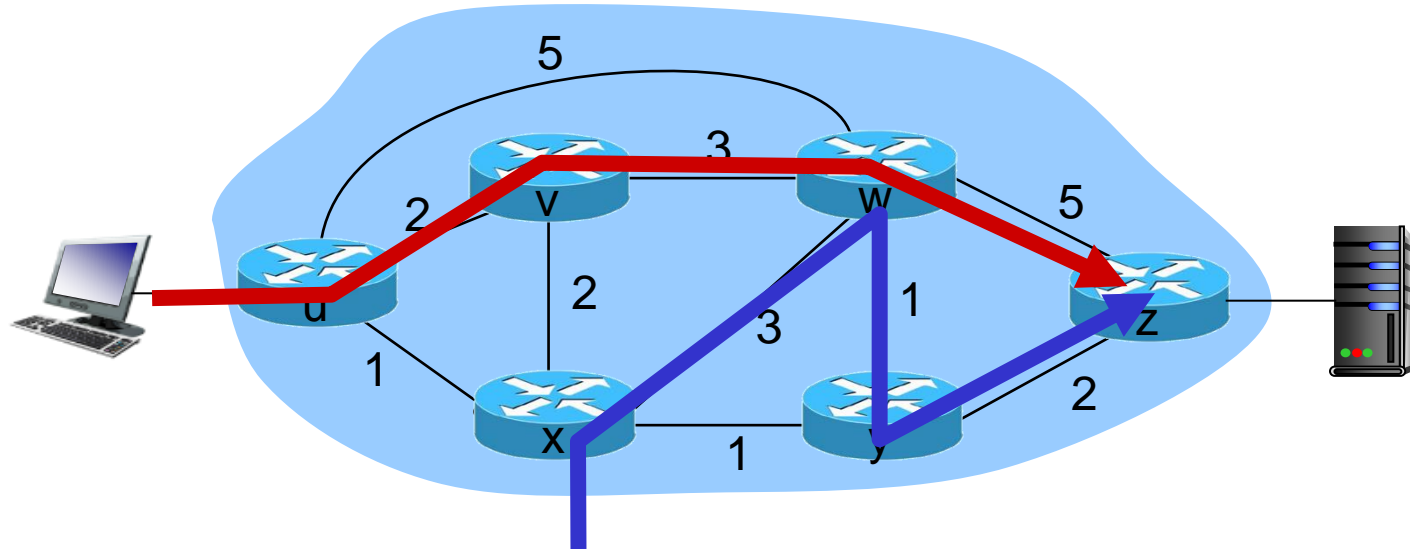
# Traffic engineering: difficult



Q: what if network operator wants to split u-to-z traffic along uvwz *and* uxyz (load balancing)?

A: can't do it (or need a new routing algorithm)

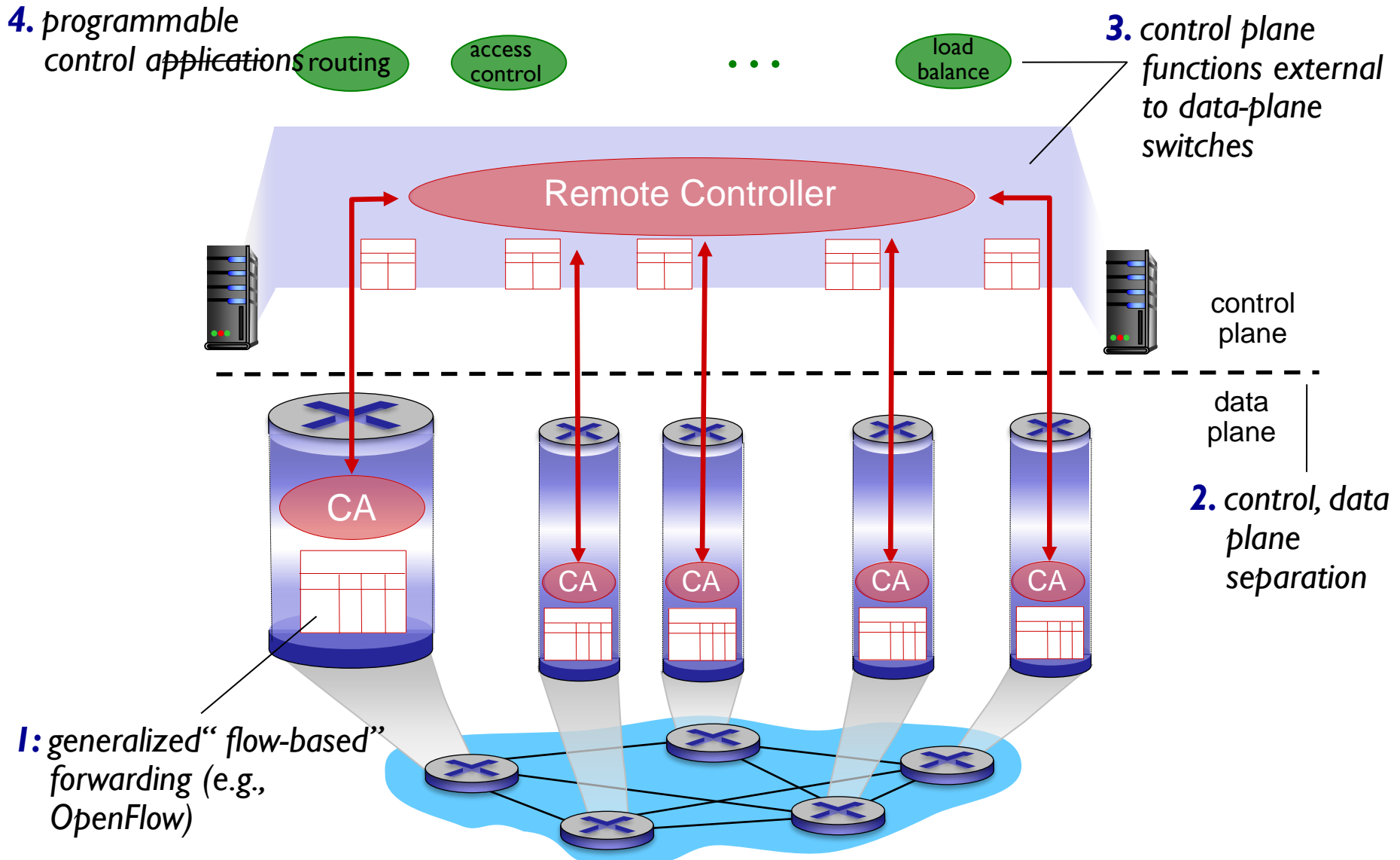
# Traffic engineering: difficult



Q: what if w wants to route blue and red traffic differently?

A: can't do it (with destination based forwarding, and LS, DV routing)

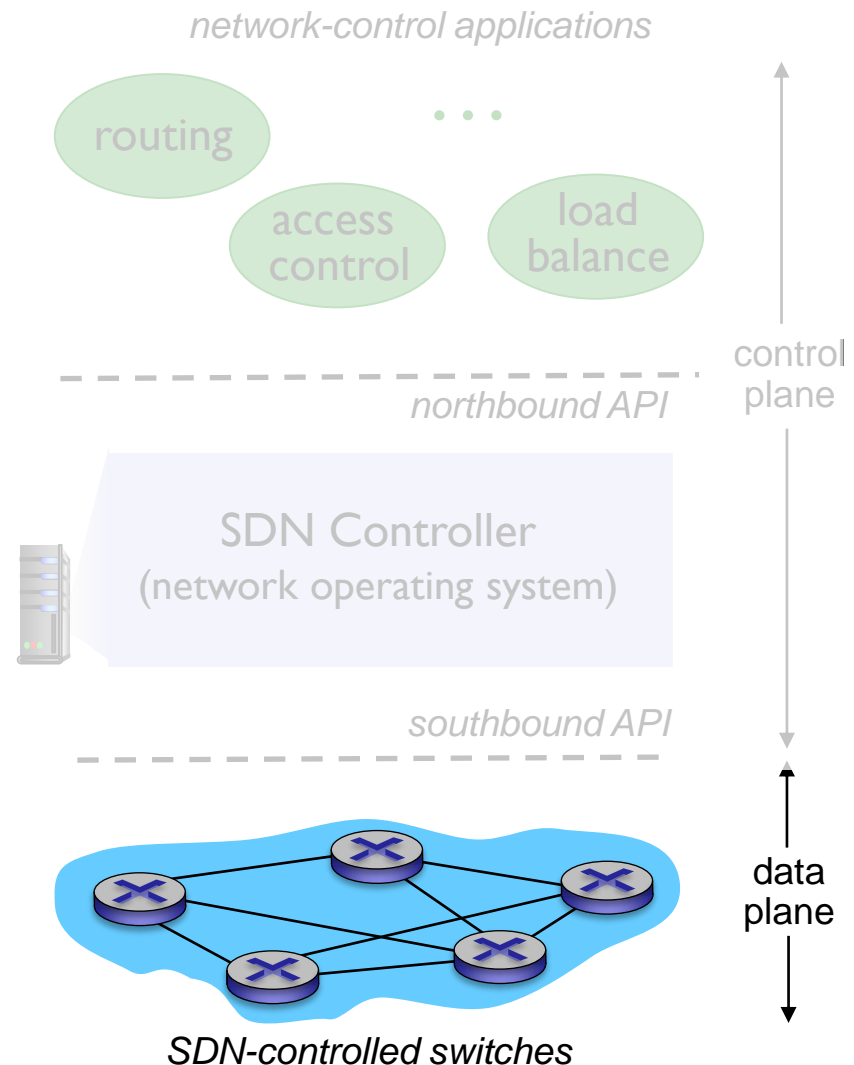
# Software defined networking (SDN)



# SDN perspective: data plane switches

## *Data plane switches*

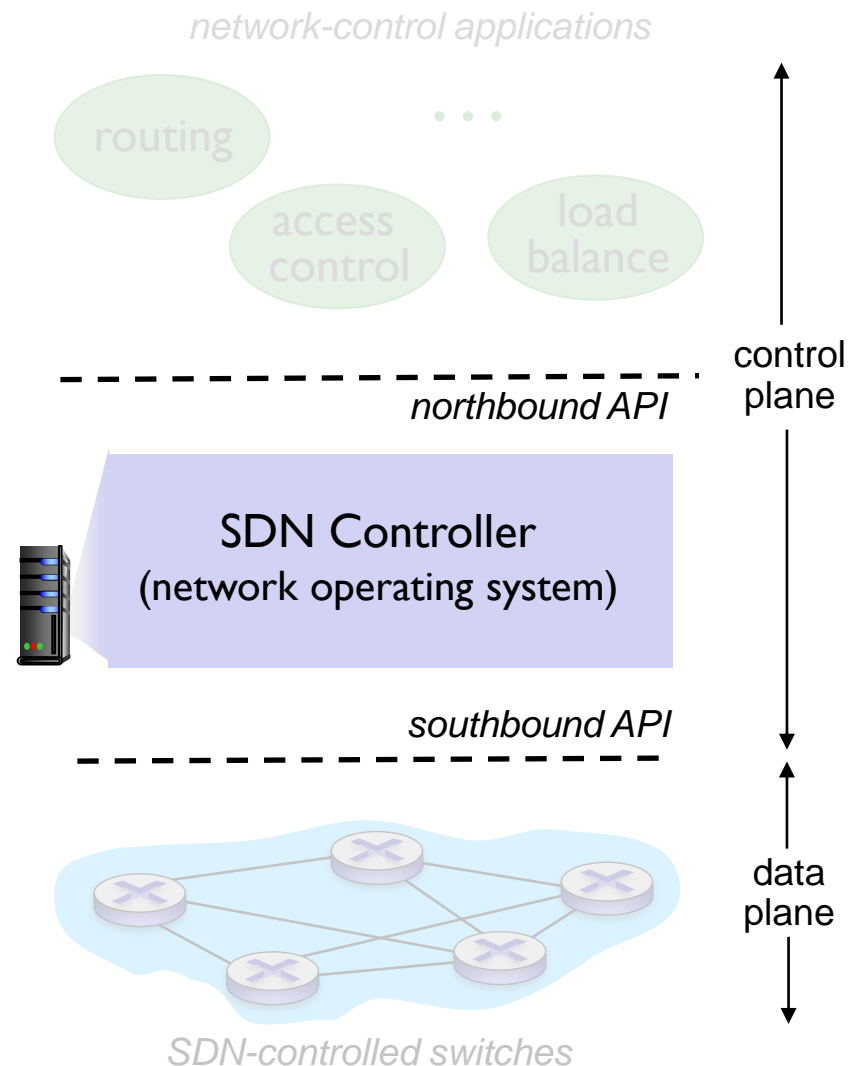
- fast, simple, commodity switches implementing generalized data-plane forwarding (Section 4.4) in hardware
- switch flow table computed, installed by controller
- API for table-based switch control (e.g., OpenFlow)
  - defines what is controllable and what is not
- protocol for communicating with controller (e.g., OpenFlow)



# SDN perspective: SDN controller

## *SDN controller (network OS):*

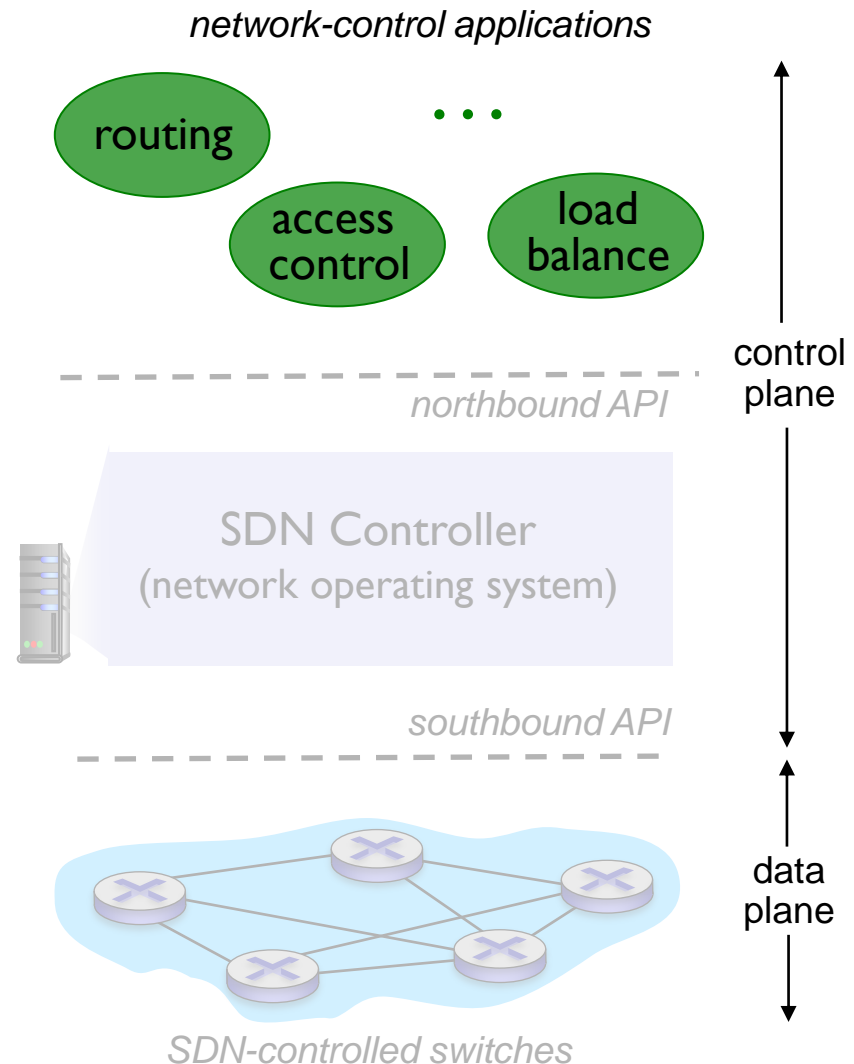
- maintain network state information
- interacts with network control applications “above” via northbound API
- interacts with network switches “below” via southbound API
- implemented as distributed system for performance, scalability, fault-tolerance, robustness



# SDN perspective: control applications

## *network-control apps:*

- “brains” of control: implement control functions using lower-level services, API provided by SDN controller
- *unbundled*: can be provided by 3<sup>rd</sup> party: distinct from routing vendor, or SDN controller



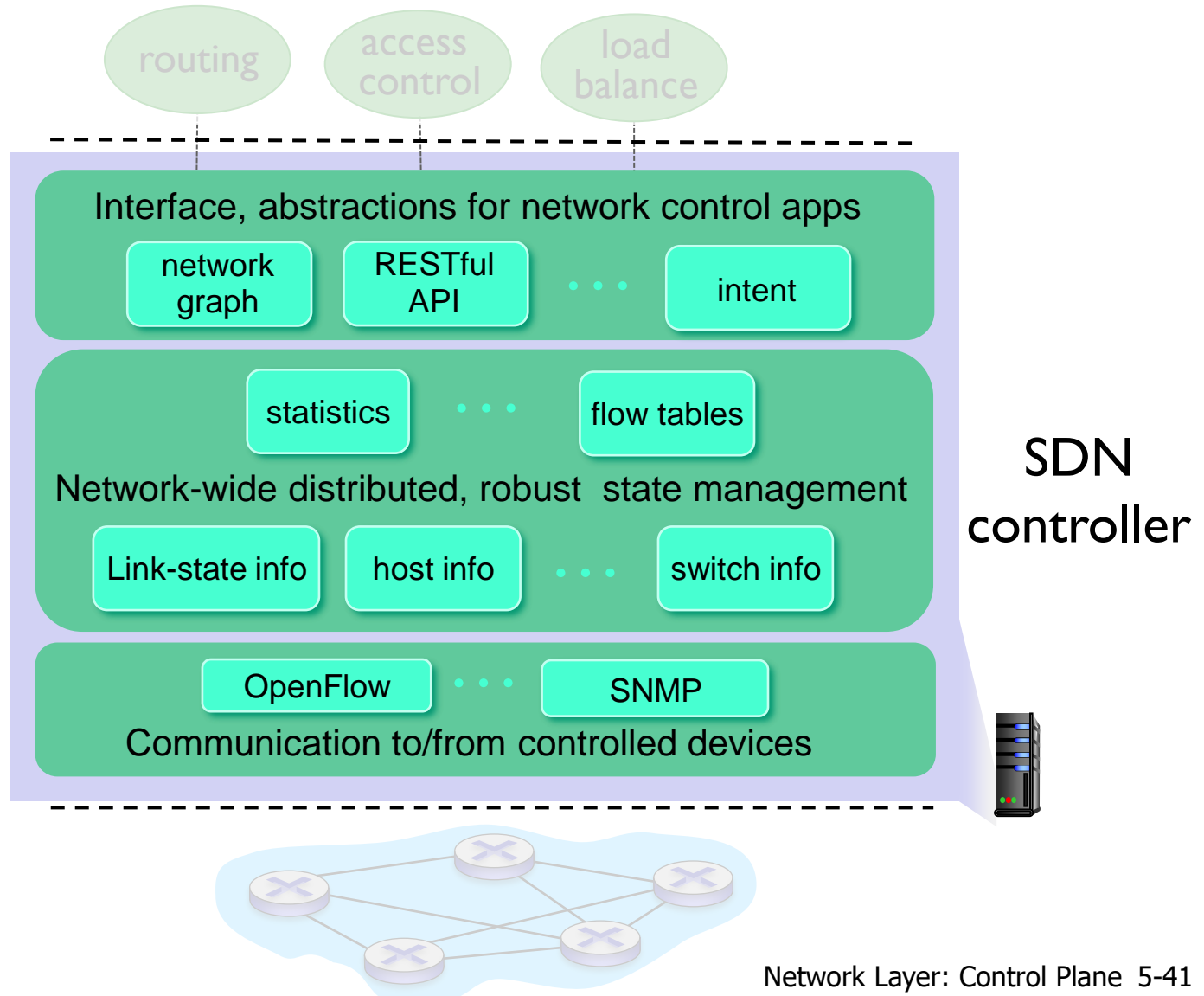


# Components of SDN controller

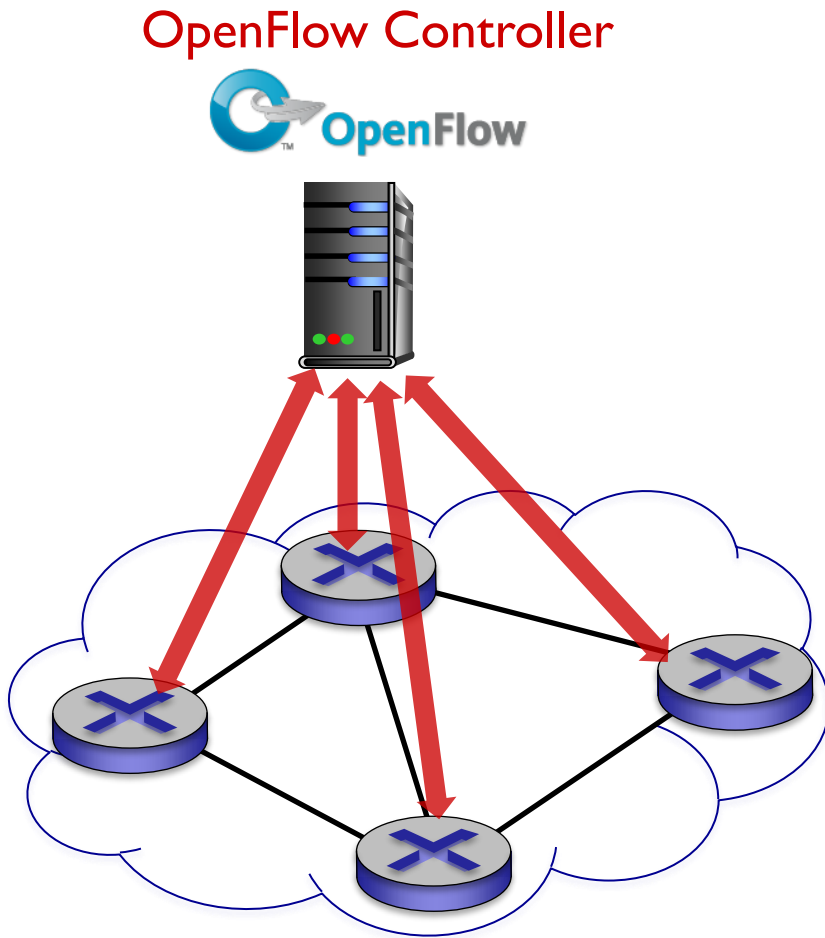
**Interface layer to network control apps:** abstractions API

**Network-wide state management layer:** state of networks links, switches, services: a *distributed database*

**communication layer:** communicate between SDN controller and controlled switches



# OpenFlow protocol

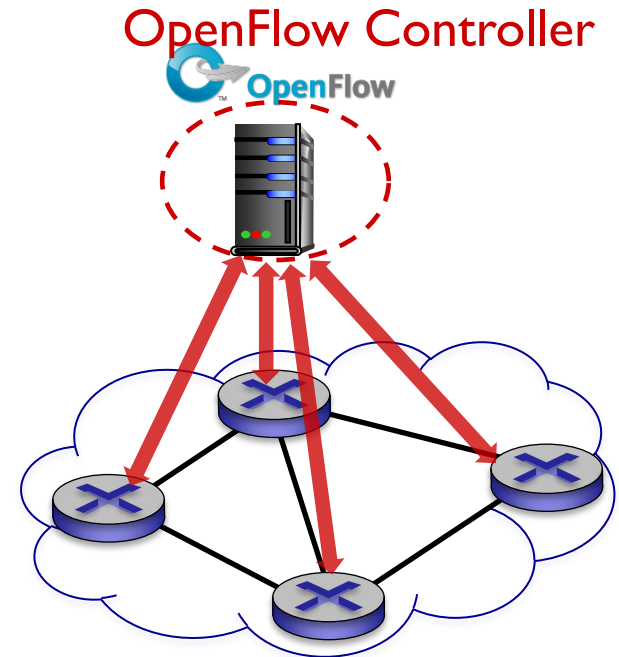


- operates between controller, switch
  - optional encryption
- TCP used to exchange messages
  - controller-to-switch
  - asynchronous (switch to controller)
  - symmetric (misc)

# OpenFlow: controller-to-switch messages

## Key controller-to-switch messages

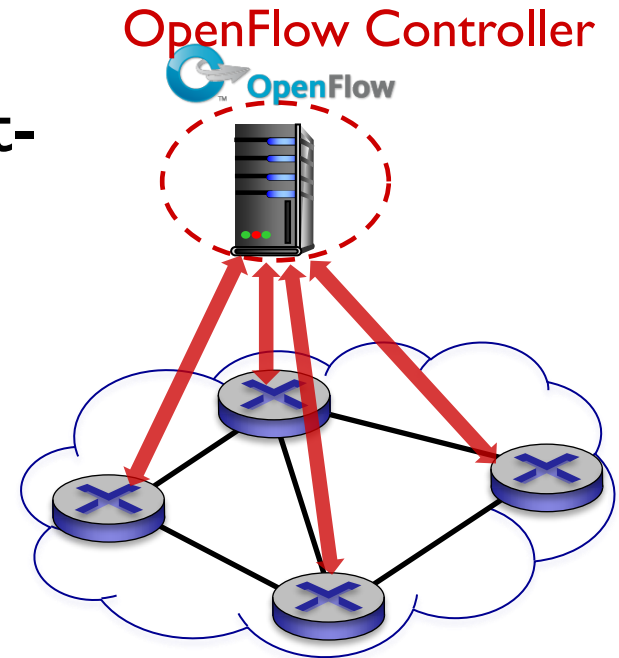
- **features:** controller queries switch features, switch replies
- **configure:** controller queries/sets switch configuration parameters
- **modify-state:** add, delete, modify flow entries in the OpenFlow tables
- **packet-out:** controller can send this packet out of specific switch port



# OpenFlow: switch-to-controller messages

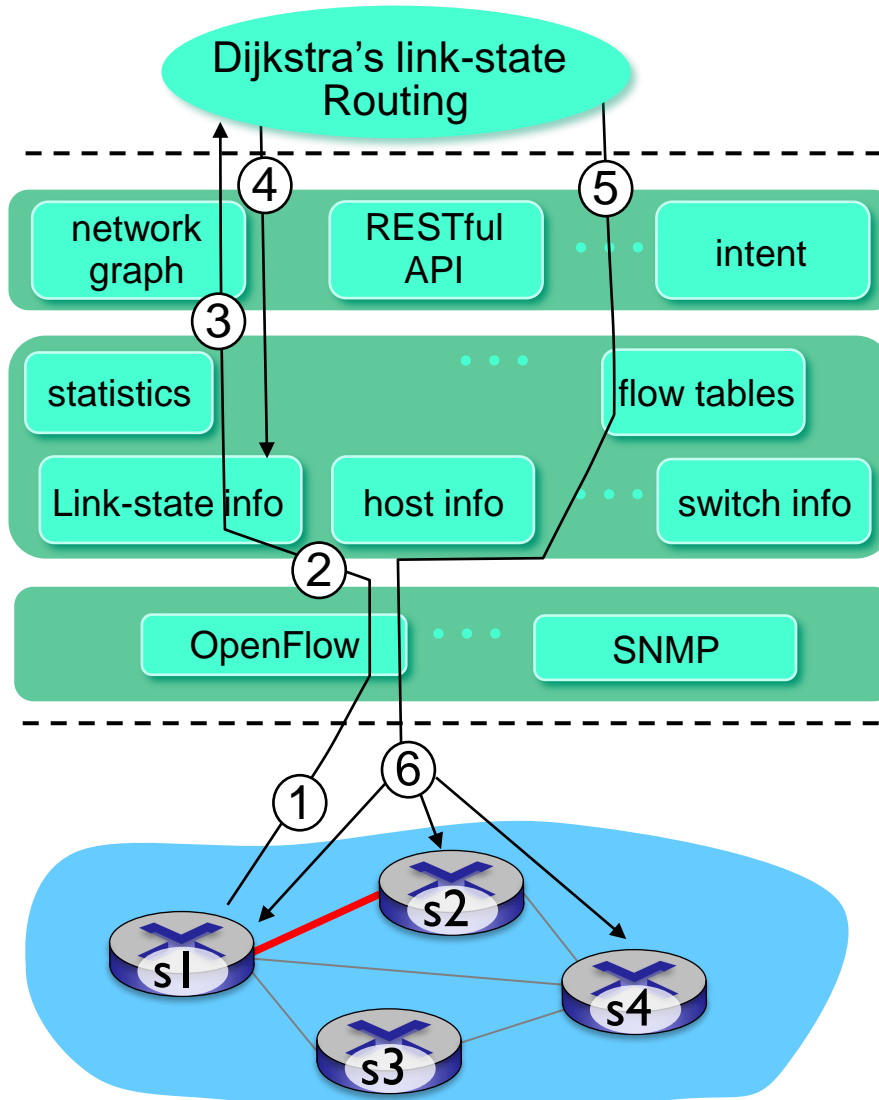
## Key switch-to-controller messages

- **packet-in:** transfer packet (and its control) to controller. See packet-out message from controller
- **flow-removed:** flow table entry deleted at switch
- **port status:** inform controller of a change on a port.



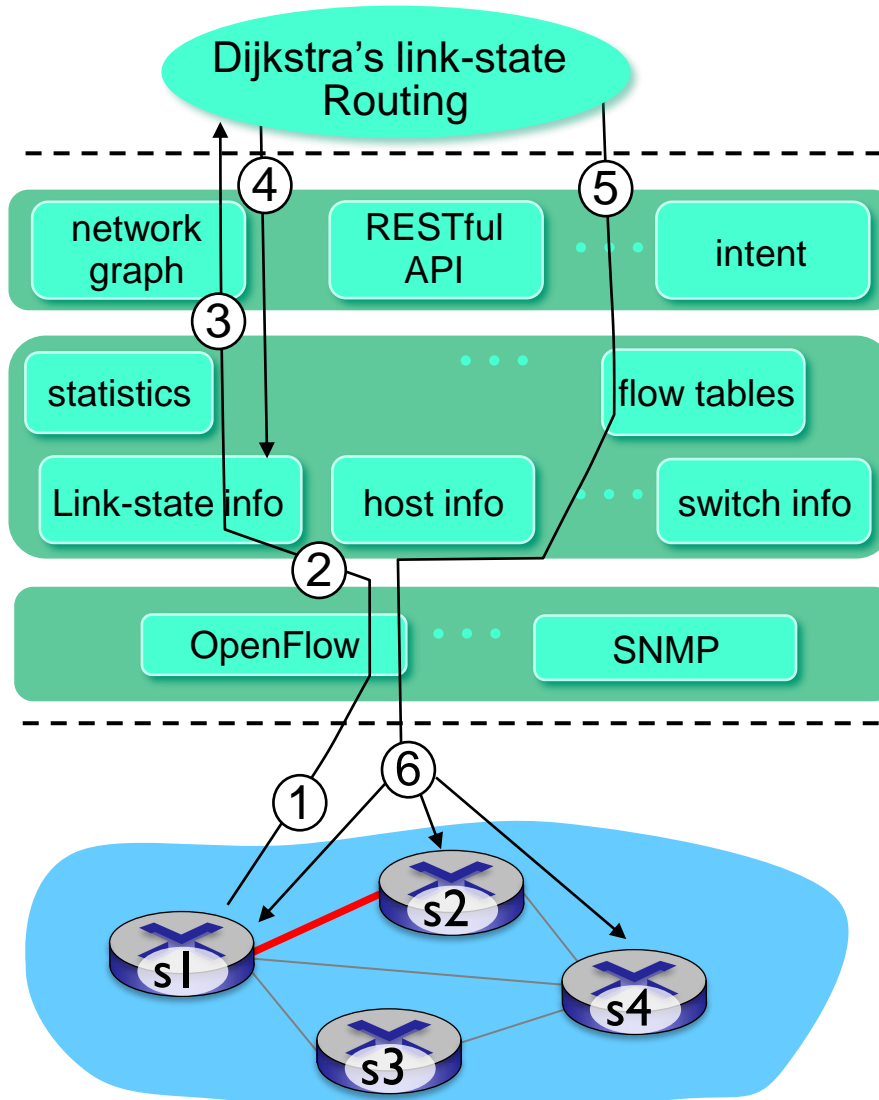
Fortunately, network operators don't “program” switches by creating/sending OpenFlow messages directly. Instead use higher-level abstraction at controller

# SDN: control/data plane interaction example



- ① SI, experiencing link failure using OpenFlow port status message to notify controller
- ② SDN controller receives OpenFlow message, updates link status info
- ③ Dijkstra's routing algorithm application has previously registered to be called when ever link status changes. It is called.
- ④ Dijkstra's routing algorithm access network graph info, link state info in controller, computes new routes

# SDN: control/data plane interaction example



- ⑤ link state routing app interacts with flow-table-computation component in SDN controller, which computes new flow tables needed
- ⑥ Controller uses OpenFlow to install new tables in switches that need updating

# Chapter 5: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the Internet: OSPF

5.4 routing among the ISPs: BGP

5.5 The SDN control plane

5.6 ICMP: The Internet Control Message Protocol

5.7 Network management and SNMP

# ICMP: internet control message protocol

- used by hosts & routers to communicate network-level information
  - error reporting: unreachable host, network, port, protocol
  - echo request/reply (used by ping)
- network-layer “above” IP:
  - ICMP msgs carried in IP datagrams
- **ICMP message:** type, code plus first 8 bytes of IP datagram causing error

<u>Type</u>	<u>Code</u>	<u>description</u>
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header



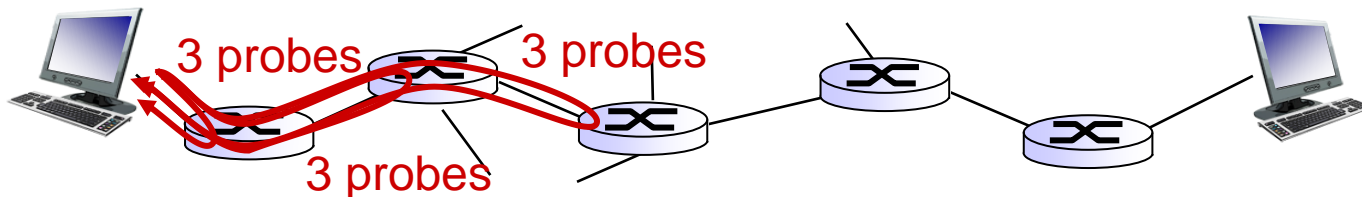
# Traceroute and ICMP

- source sends series of UDP segments to destination
  - first set has TTL = 1
  - second set has TTL=2, etc.
  - unlikely port number
- when datagram in  $n$ th set arrives to  $n$ th router:
  - router discards datagram and sends source ICMP message (type 11, code 0)
  - ICMP message include name of router & IP address

- when ICMP message arrives, source records RTTs

## *stopping criteria:*

- UDP segment eventually arrives at destination host
- destination returns ICMP “port unreachable” message (type 3, code 3)
- source stops



# Chapter 5: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the Internet: OSPF

5.4 routing among the ISPs: BGP

5.5 The SDN control plane

5.6 ICMP: The Internet Control Message Protocol

5.7 Network management and SNMP

# What is network management?

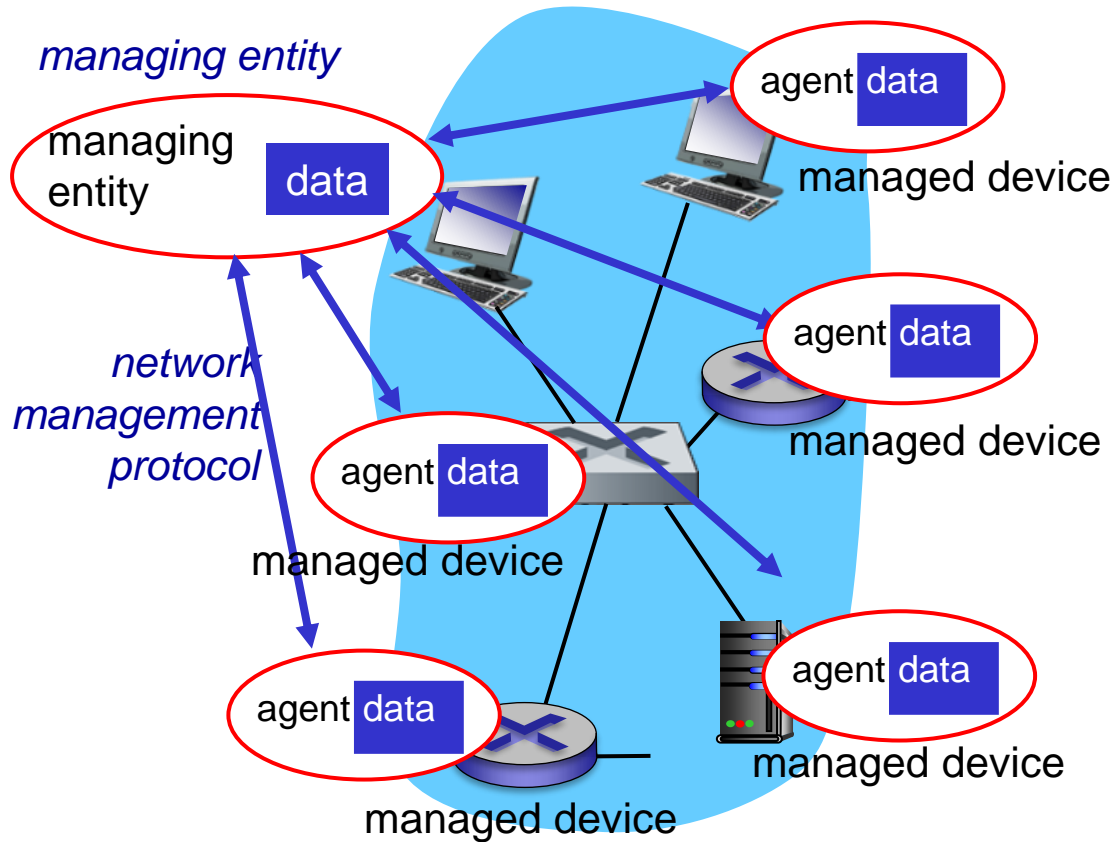
- **autonomous systems (aka “network”)**: 1000s of interacting hardware/software components
- other complex systems requiring monitoring, control:
  - jet airplane
  - nuclear power plant
  - others?



"**Network management** includes the deployment, integration and coordination of the hardware, software, and human elements to monitor, test, poll, configure, analyze, evaluate, and control the network and element resources to meet the real-time, operational performance, and Quality of Service requirements at a reasonable cost."

# Infrastructure for network management

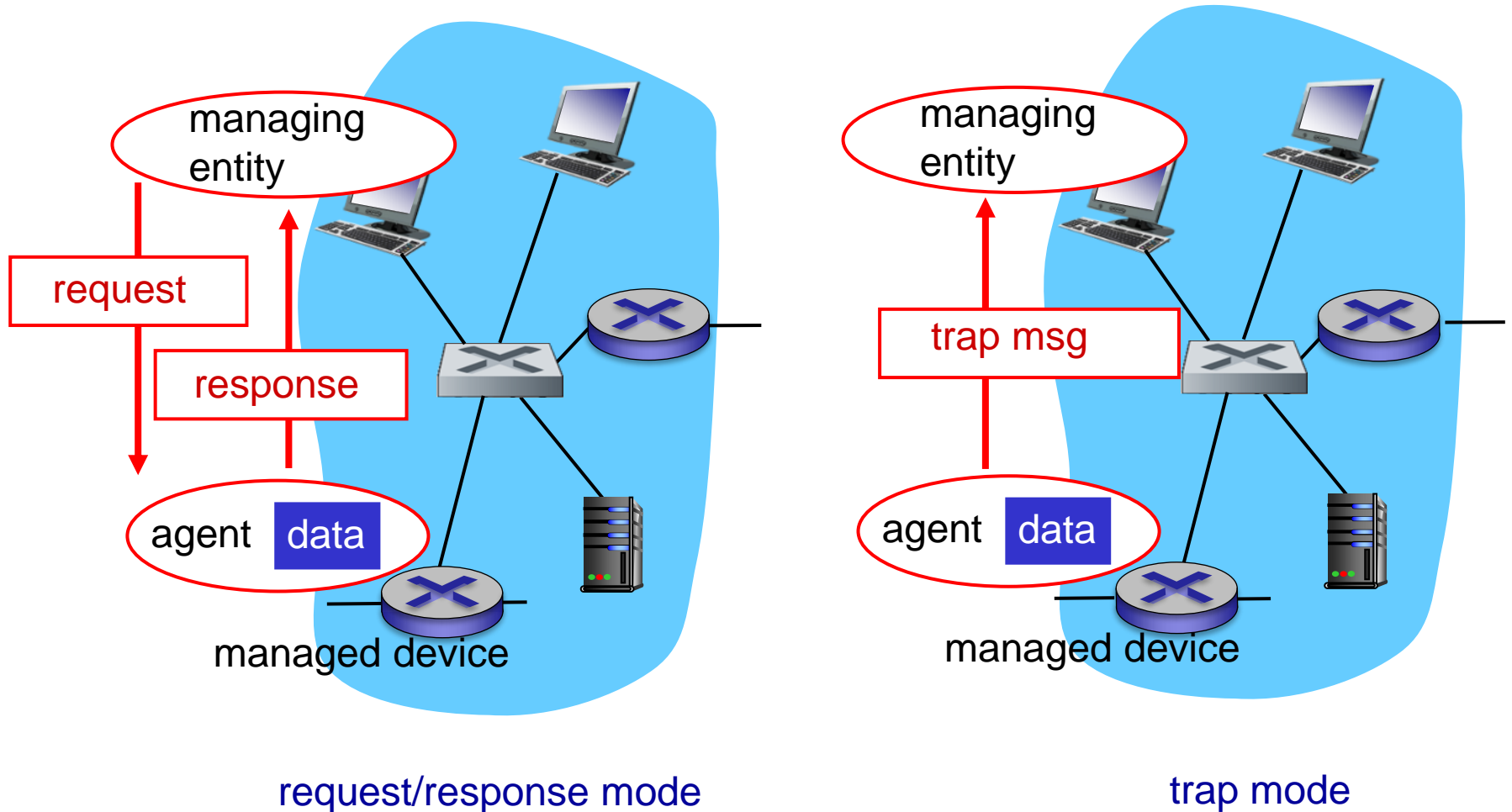
definitions:



*managed devices*  
contain *managed objects* whose data is gathered into a **Management Information Base (MIB)**

# SNMP protocol

Two ways to convey MIB info, commands:



# SNMP protocol: message types

## Message type

## Function

GetRequest  
GetNextRequest  
GetBulkRequest

manager-to-agent: “get me data”  
(data instance, next data in list, block of data)

InformRequest

manager-to-manager: here's MIB value

SetRequest

manager-to-agent: set MIB value

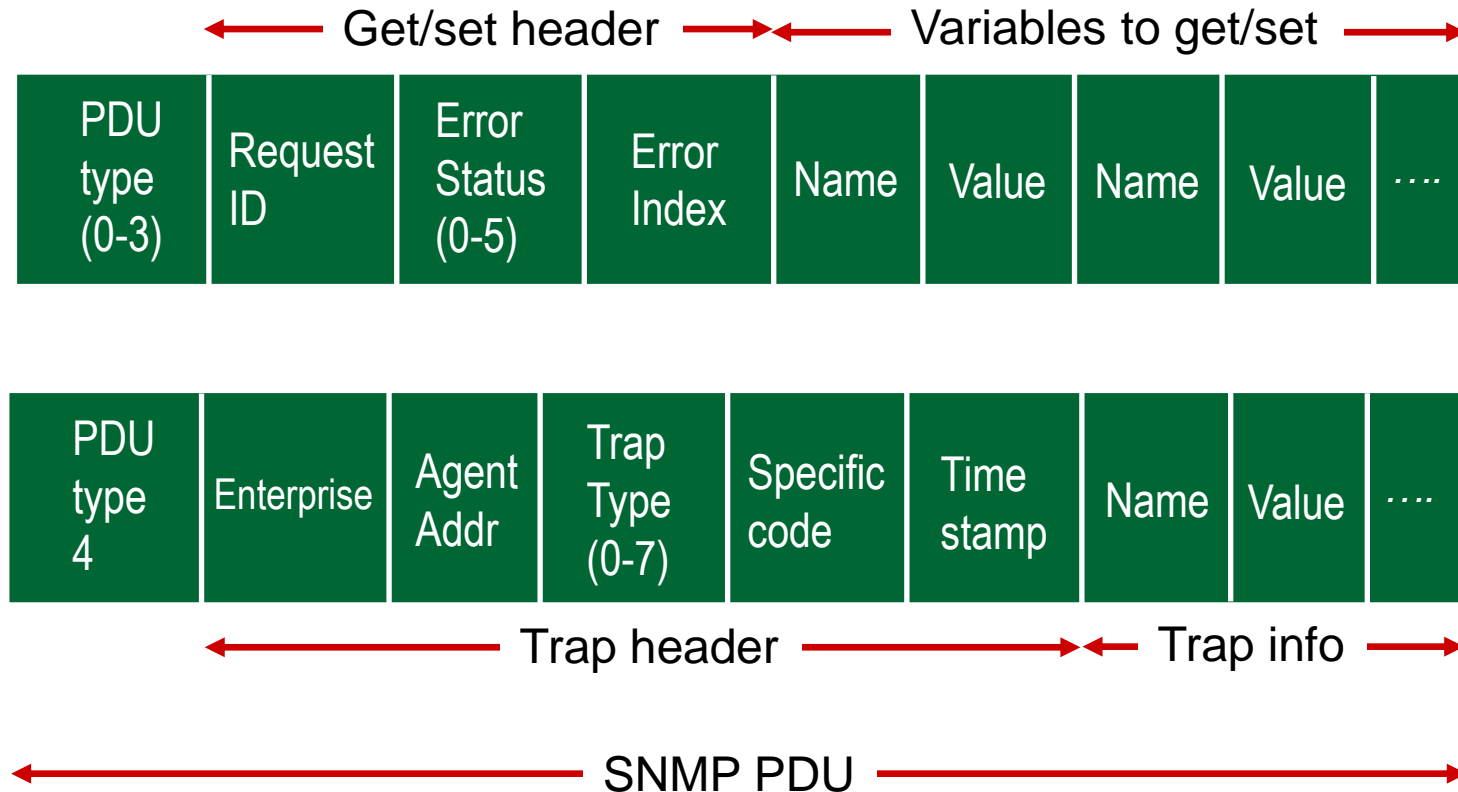
Response

Agent-to-manager: value, response to Request

Trap

Agent-to-manager: inform manager of exceptional event

# SNMP protocol: message formats



*More on network management: see earlier editions of text!*

# Chapter 5: summary

*we've learned a lot!*

- approaches to network control plane
  - per-router control (traditional)
  - logically centralized control (software defined networking)
- traditional routing algorithms
  - implementation in Internet: OSPF, BGP
- SDN controllers
  - implementation in practice: ODL, ONOS
- Internet Control Message Protocol
- network management

*next stop: link layer!*