# Internet Protocols EBU5403
# The Network Layer (Part II)
# C3

Michael Chai (michael.chai@qmul.ac.uk)
Richard Clegg (r.clegg@qmul.ac.uk)
Cunhua Pan (c.pan@qmul.ac.uk)

| | Part 1 | Part 2 | Part 3 | Part 4 |
|---|---|---|---|---|
| Ecommerce + Telecoms 1 | Richard Clegg | | Cunhua Pan | |
| Telecoms 2 | Michael Chai | | | |

# Network Control Plane: 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

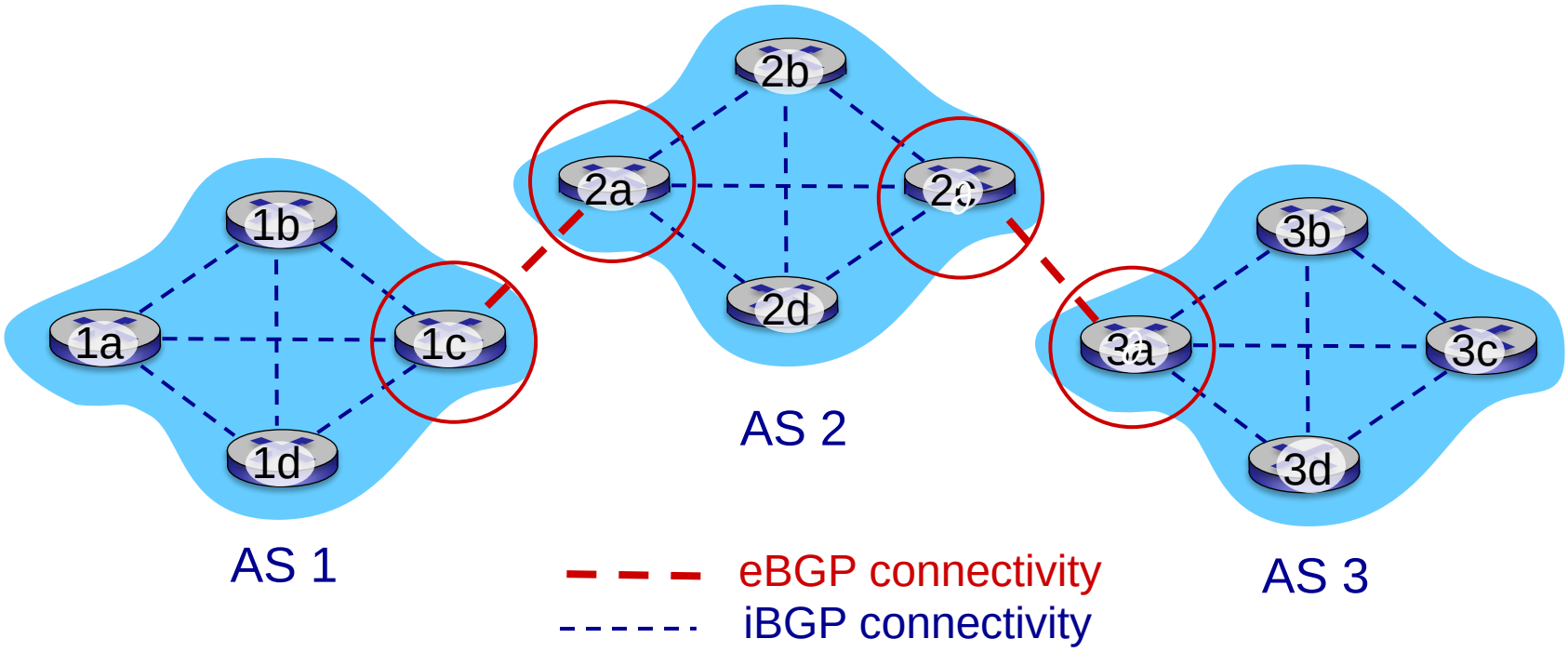
5.5 The SDN control plane

5.6 ICMP: The Internet Control Message Protocol

# Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- BGP provides each AS a means to:
  - eBGP: obtain subnet reachability information from neighboring ASes
  - iBGP: propagate reachability information to all AS-internal routers.
  - 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

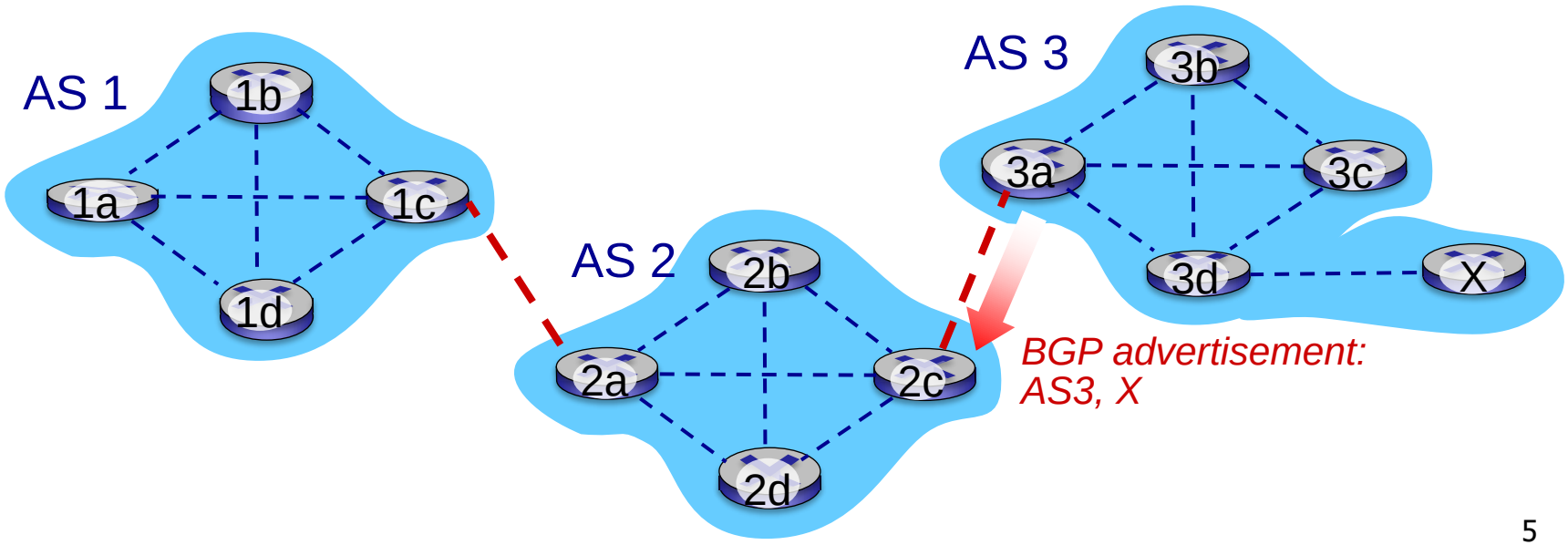




gateway routers run both eBGP and iBGP protools

# BGP basics

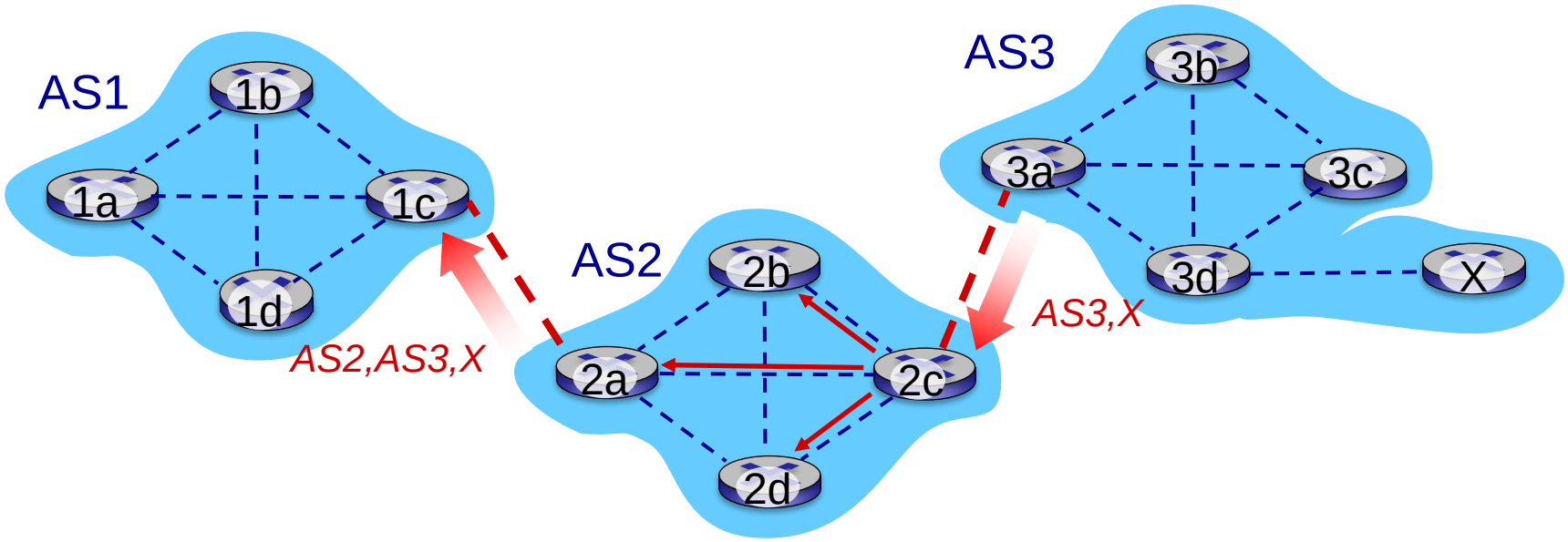
- BGP session: two BGP routers ("peers") exchange BGP messages over semi-permanent TCP connection:
  - advertising *paths* to different destination network prefixes (BGP is a "path vector" protocol)
- when AS3 gateway router 3a advertises path AS3,X to AS2 gateway router 2c:
  - AS3 *promises* to AS2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



# Path attributes and BGP routes

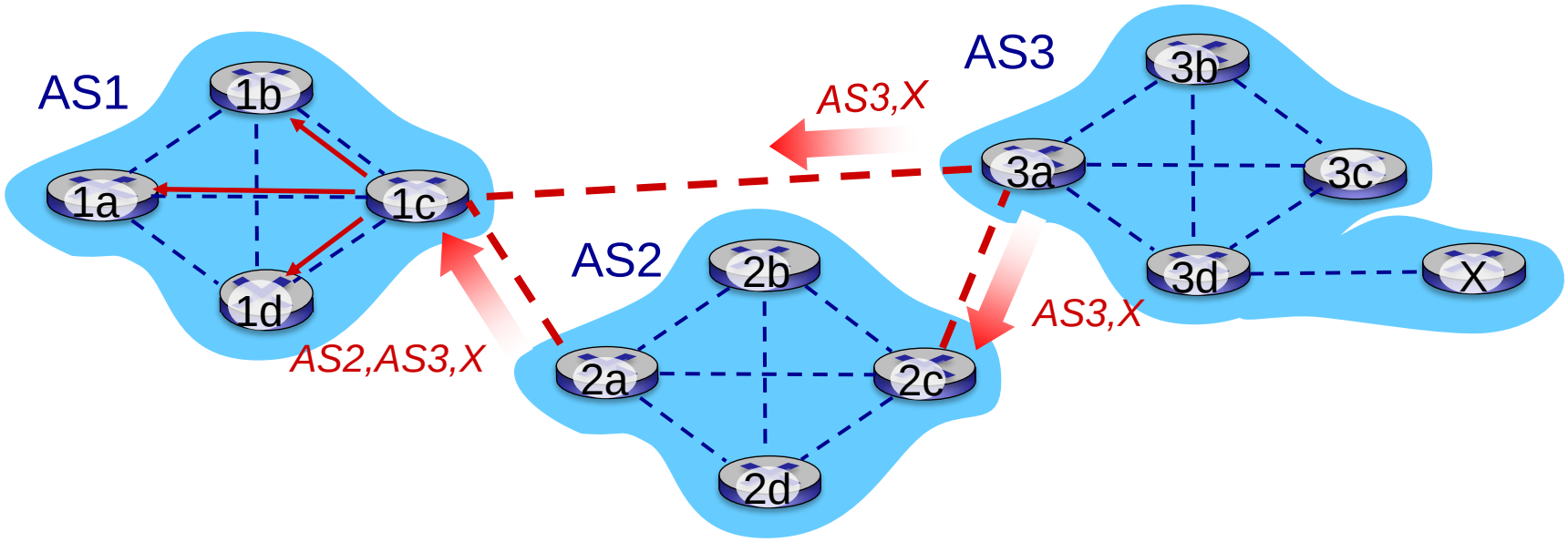
- advertised prefix includes BGP attributes
  - prefix + attributes = "route"
- two important attributes:
  - AS-PATH: list of ASes through which prefix advertisement has passed
  - NEXT-HOP: indicates specific internal-AS router to next-hop AS
- *Policy-based routing:*
  - gateway receiving route advertisement uses *import policy* to accept/decline path (e.g., never route through AS Y).
  - AS policy also determines whether to *advertise* path to other 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, AS1 gateway router 1c chooses path *AS3,X, and advertises path within AS1 via iBGP*

# 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

# Test your understanding

BGP trades…

A. Distance information

B. Link state information

C. Reachability information

D. All the provided answers

# Network Control Plane: outline

5.1 introduction

5.2 routing protocols

- link state
- distance vector

5.3 intra-AS routing in the Internet: OSPF

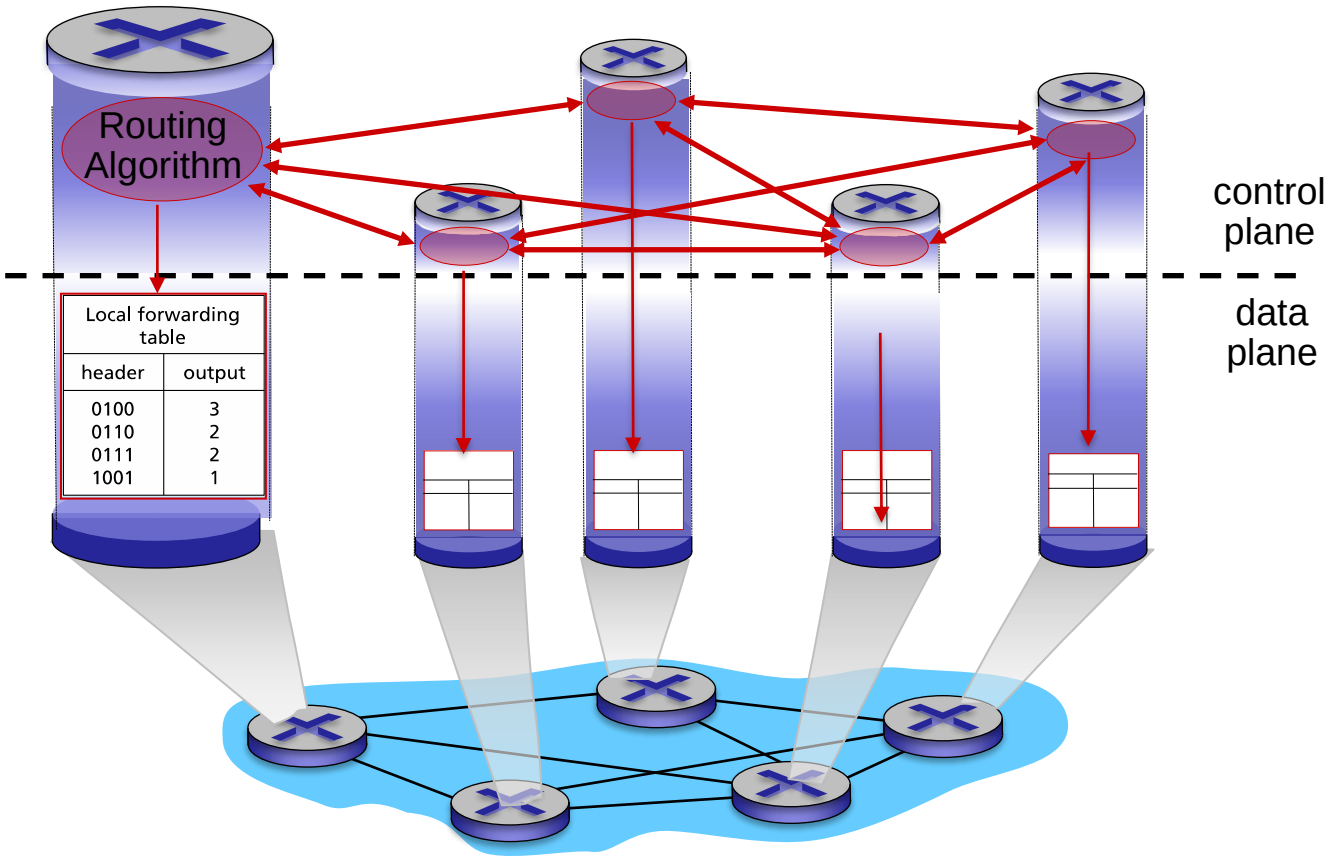
5.4 routing among the ISPs: BGP

5.5 The SDN control plane

5.6 ICMP: The Internet Control Message Protocol

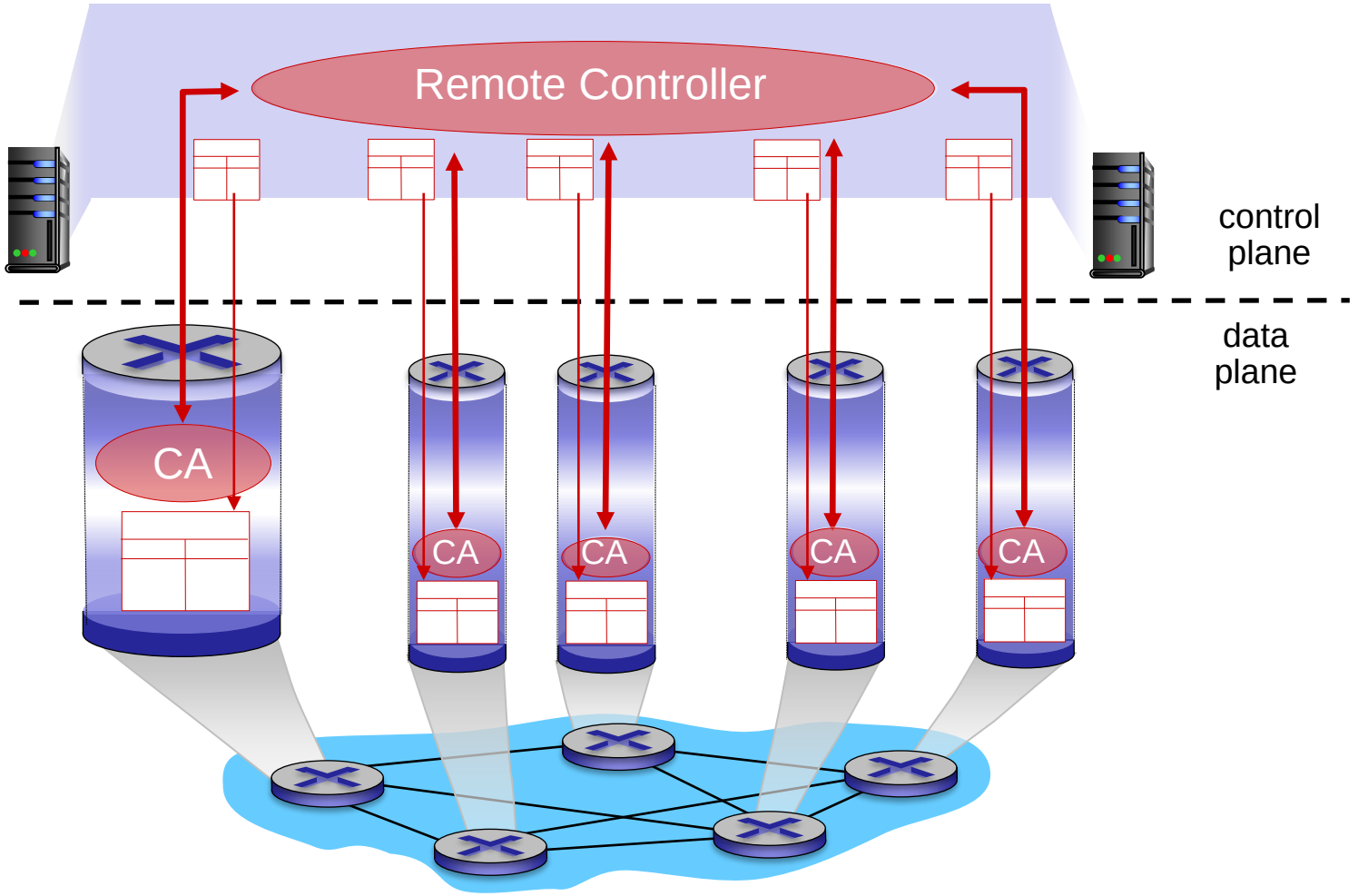
# Recall: per-router control plane

Individual routing algorithm components *in each and every router* interact with each other in control plane to compute forwarding tables



# Recall: logically centralized control plane

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

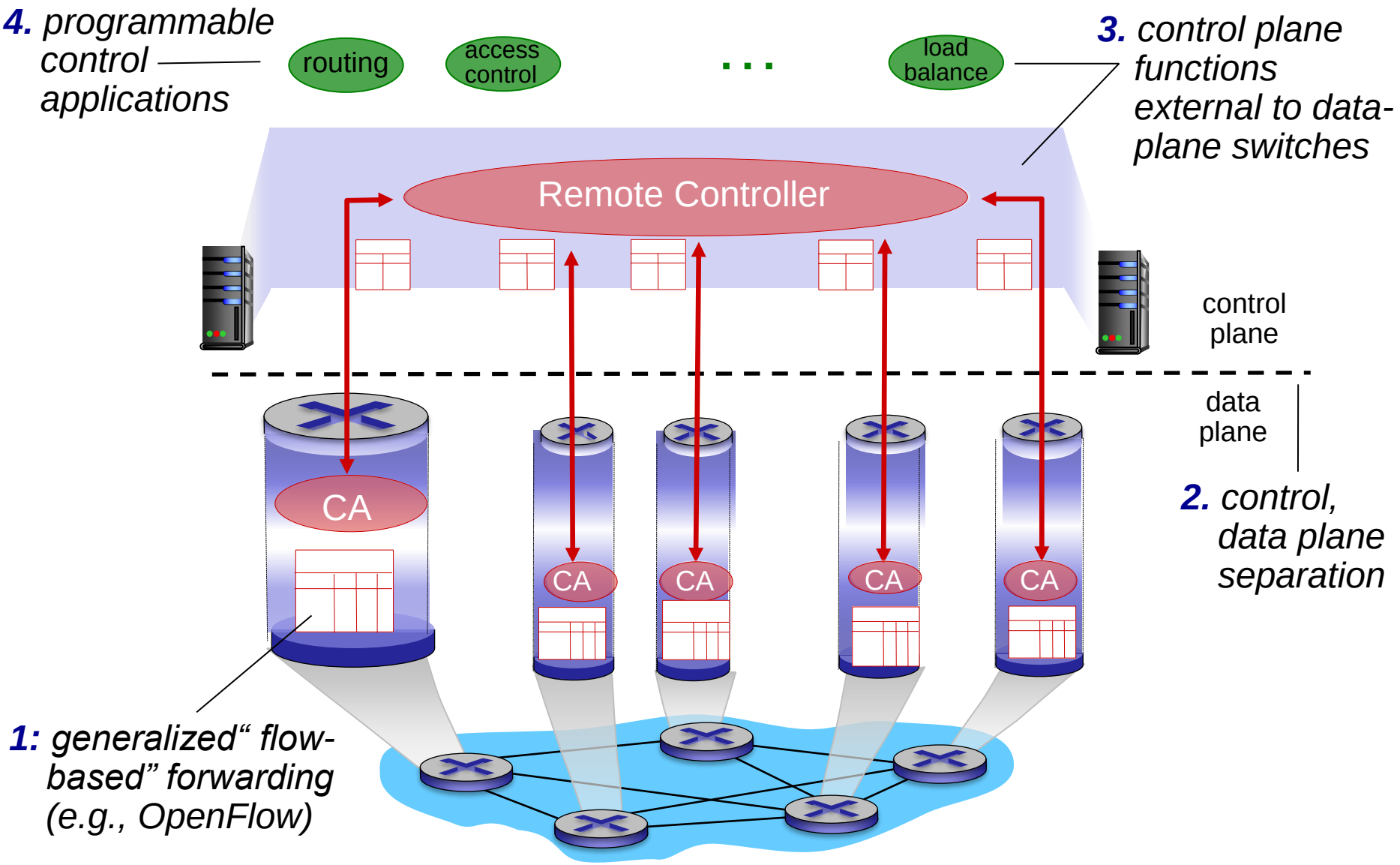


# Software defined networking (SDN)

*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

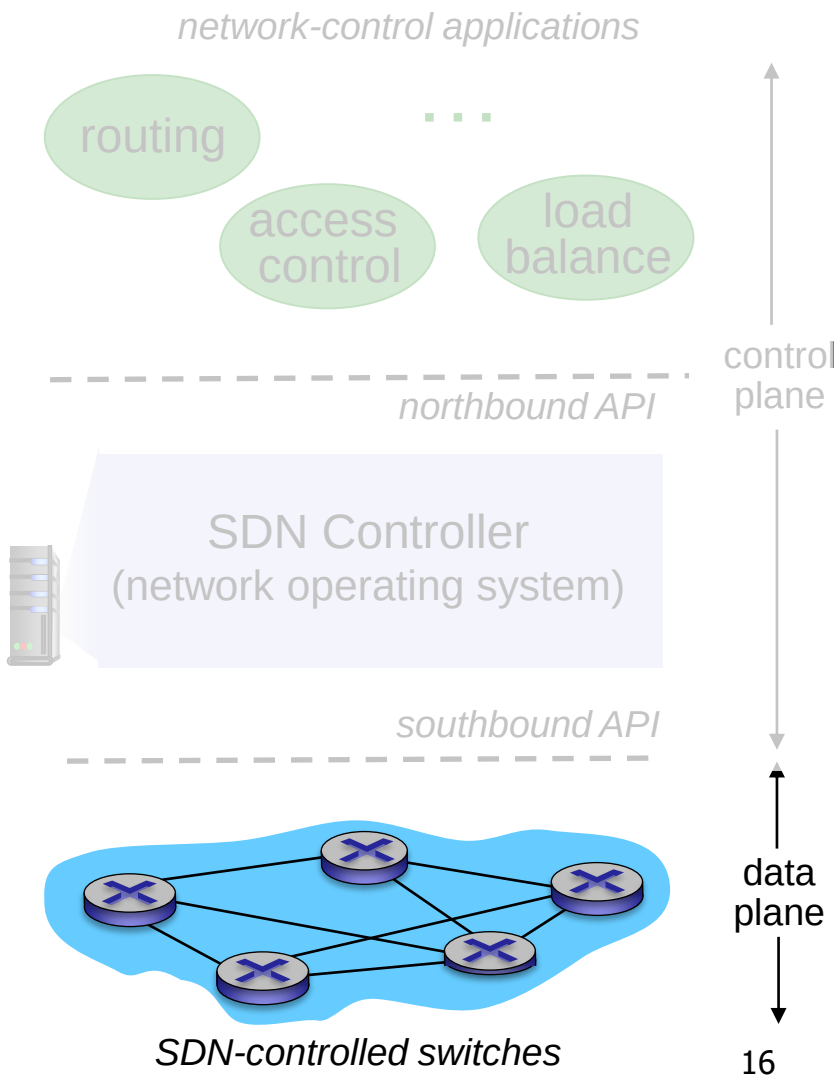
# Software defined networking (SDN)



# SDN perspective: data plane switches

## *Data plane switches*

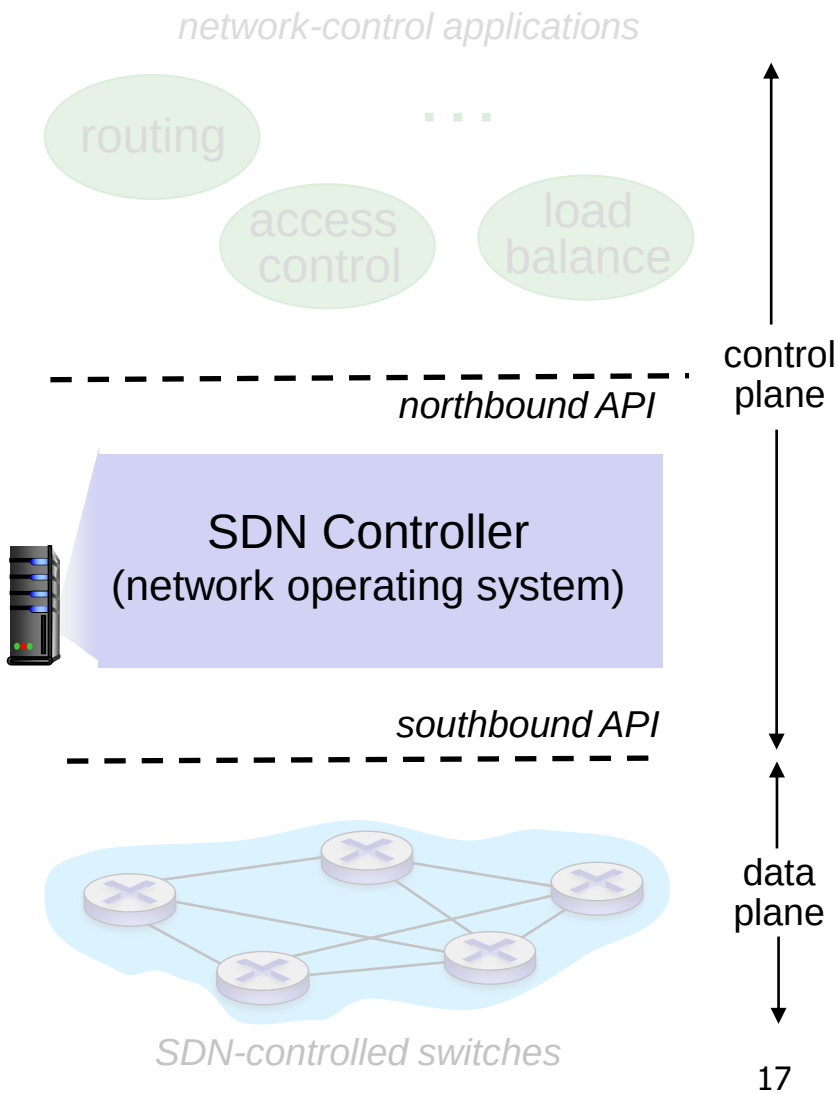
- fast, simple, commodity switches implementing generalized data-plane forwarding in hardware
- switch flow table computed, installed by controller
- API for table-based switch control (e.g., OpenFlow)
  - defines what is controllable and what is not
- protocol for communicating with controller (e.g.,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



# SDN perspective: SDN controller

*SDN controller (network OS):*

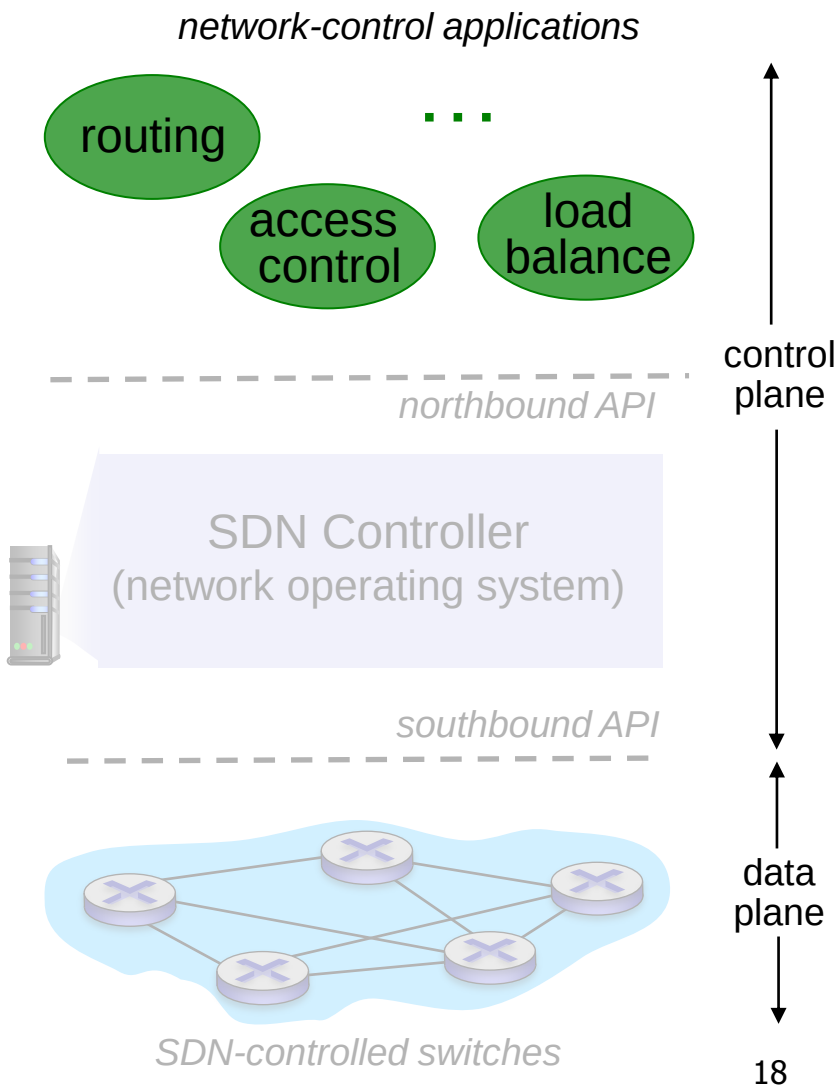
- maintain network state information
- interacts with network control applications "above" via northbound API
- interacts with network switches "below" via southbound API
- implemented as distributed system for performance, scalability, fault-tolerance, robustness
- southbound protocol – connects controller to switch.
- northbound protocol -- connects controller to apps



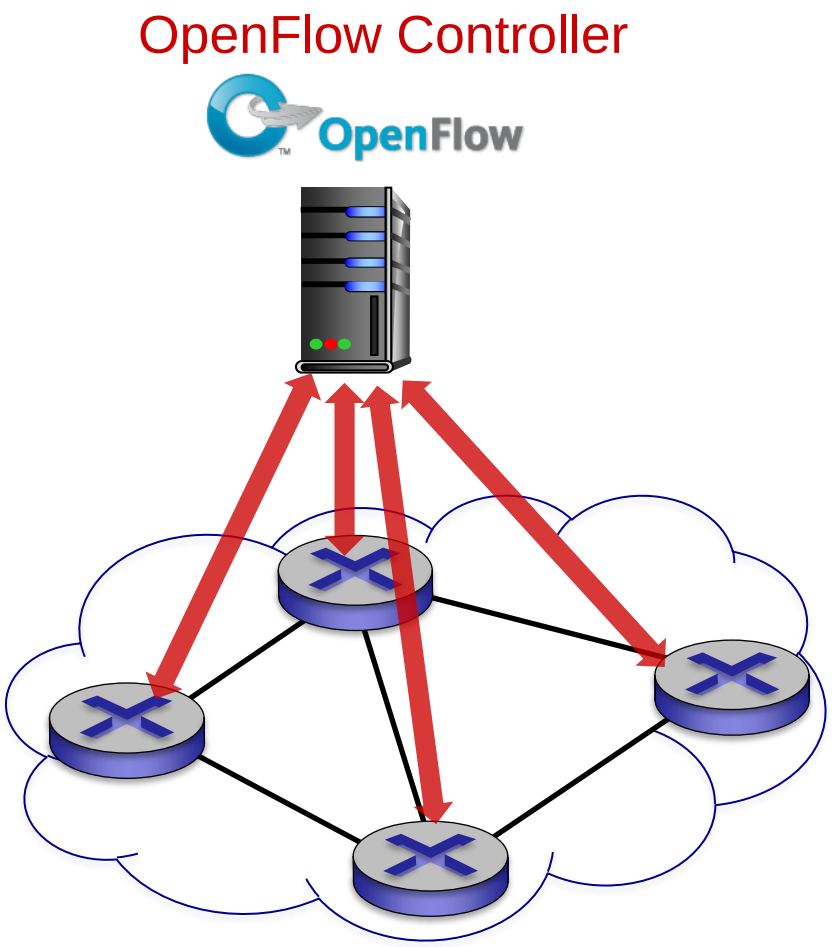
# SDN perspective: control applications

*network-control apps:*

- "brains" of control: implement control functions using lower-level services, API provided by SDN controller
- *unbundled:* applications can be written by anyone, not just company who sold switch or company who created controller

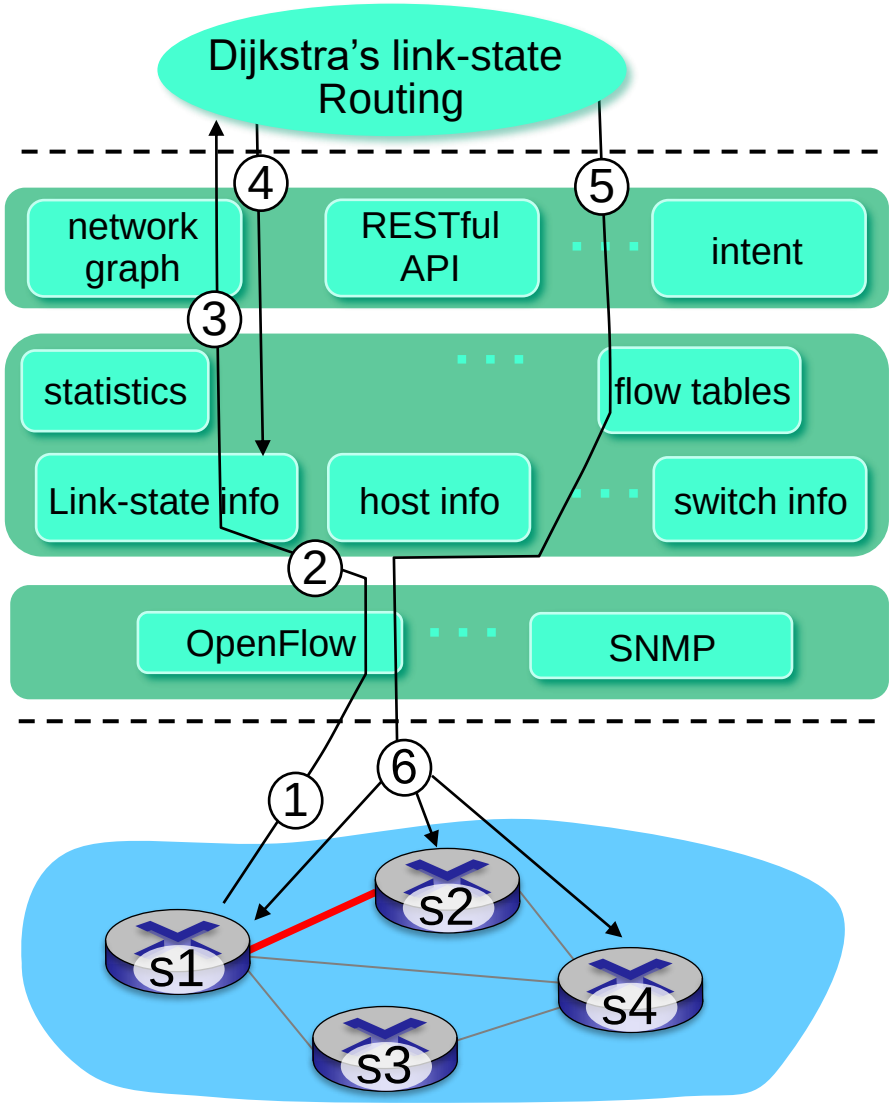


# OpenFlow protocol




- operates between controller, switch
- TCP used to exchange messages
  - optional encryption
- three classes of OpenFlow messages:
  - controller-to-switch
  - asynchronous (switch to controller)
  - symmetric (misc)

# SDN: control/data plane interaction example



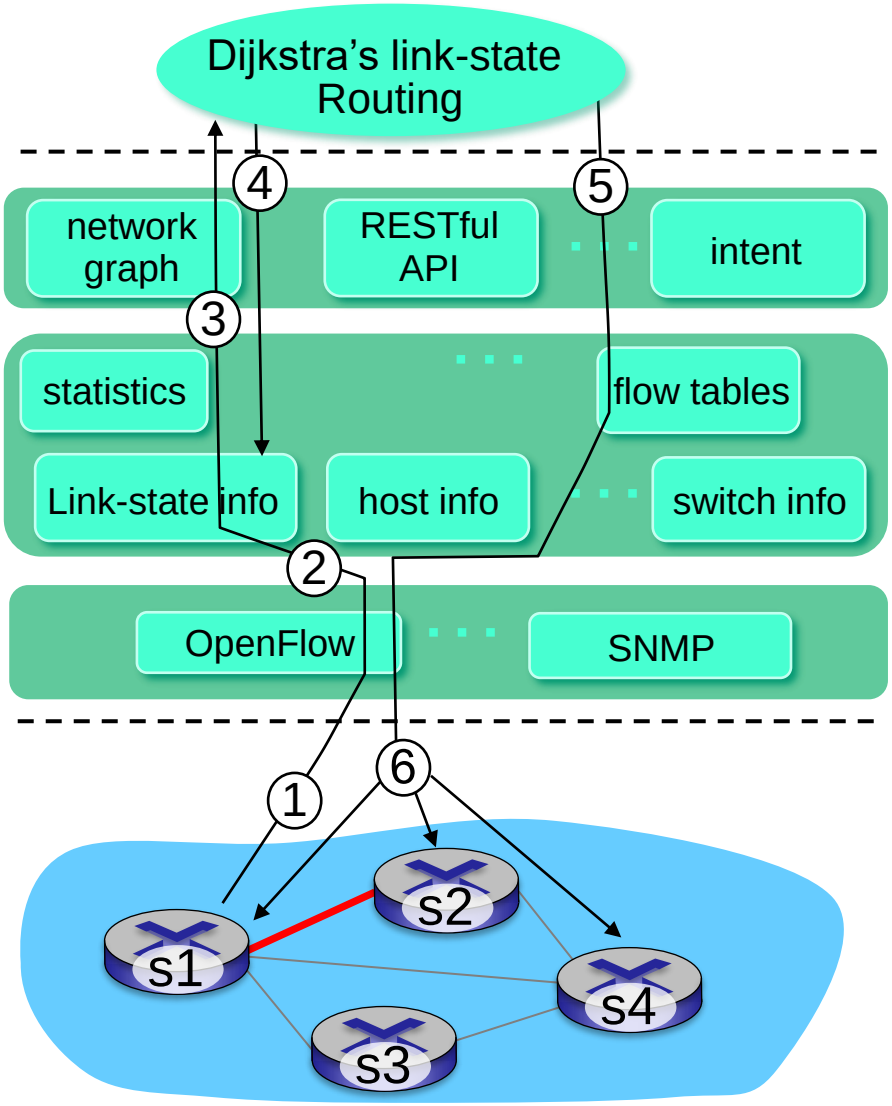

① S1, experiencing link failure using OpenFlow port status message to notify controller

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

# Test your understanding

Briefly describe the differences between the routing using match-action in Open Flow and the traditional routing using a link-state protocol.

# Test your understanding

- Traditional routing
  - Dijkstra algorithm distributed between routers
  - Forwarding table traditionally on individual router
- Openflow routing
  - Logically centralised in controller
  - Match-action table
  - Open Flow centralised in controller

# Network Control Plane: 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

# 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

| Type | Code | description |
|---|---|---|
| 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

# Traceroute example

Tracing route to www.bupt.edu.cn
[124.127.207.2]
(from QMUL)

In QMUL

| | | | | |
|---|---|---|---|---|
| 1 | 2 ms | 2 ms | 2 ms | 161.23.60.2 |
| 2 | 3 ms | 3 ms | 3 ms | 172.23.22.17 |
| 3 | 2 ms | 2 ms | 13 ms | 172.23.48.194 |
| 4 | 2 ms | 2 ms | 2 ms | 172.23.56.1 |
| 5 | 3 ms | 3 ms | 3 ms | 172.23.8.14 |
| 6 | 3 ms | 2 ms | 2 ms | 172.23.8.18 |
| 7 | 2 ms | 2 ms | 2 ms | 172.23.56.10 |
| 8 | 3 ms | 2 ms | 2 ms | 172.23.52.17 |
| 9 | 2 ms | 2 ms | 4 ms | 172.23.16.162 |
| 10 | 3 ms | 3 ms | 3 ms | 146.97.143.217 |
| 11 | 3 ms | 2 ms | 3 ms | 146.97.35.233 |
| 12 | 4 ms | 4 ms | 3 ms | 146.97.33.1 |
| 13 | 3 ms | 3 ms | 3 ms | 146.97.35.206 |

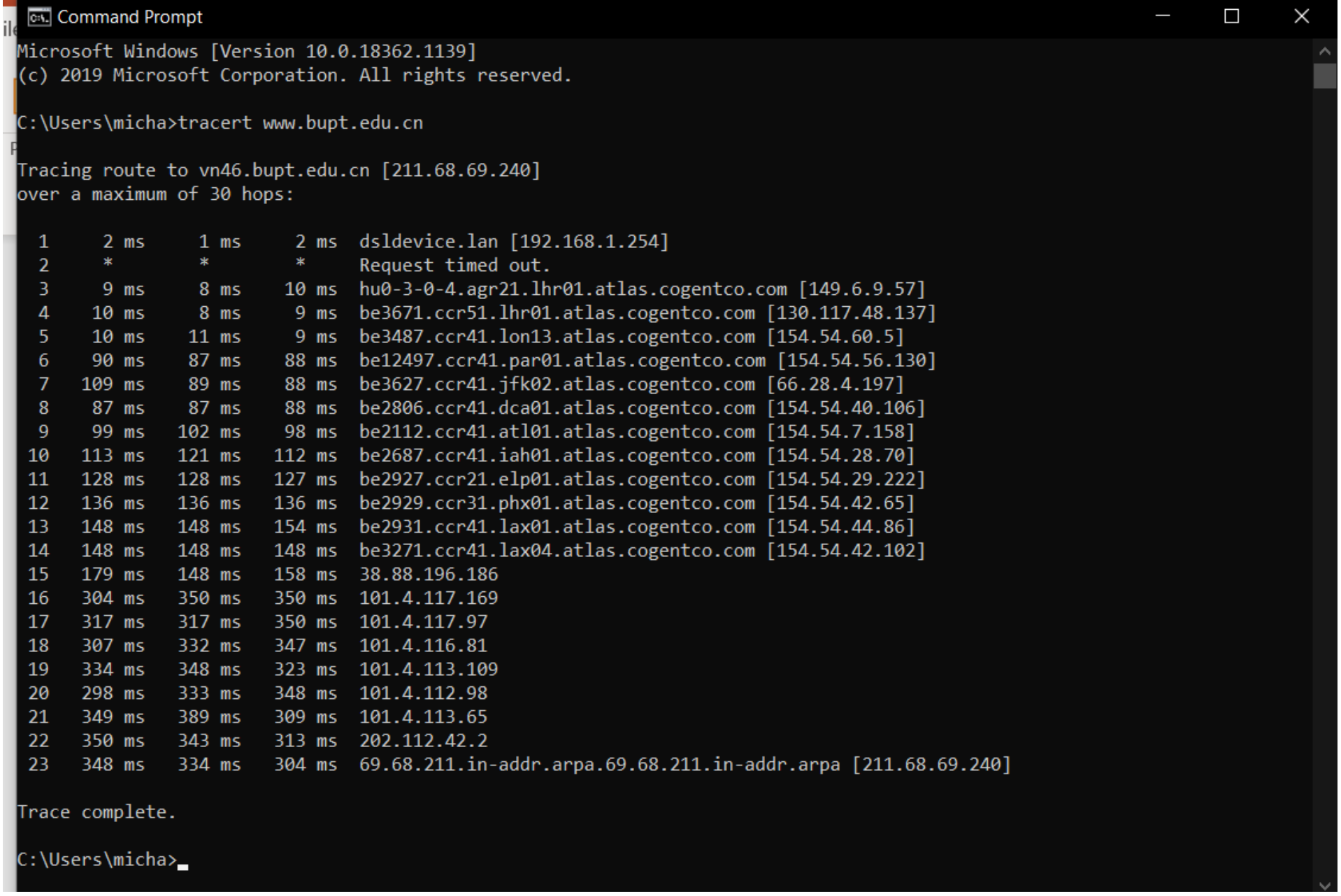
UK JANET (Joint Academic Network)
London

CHINANET (first hop is UK end of Connection)

| | | | | |
|---|---|---|---|---|
| 14 | 6 ms | 5 ms | 4 ms | 202.97.52.97 |
| 15 | 189 ms | 186 ms | 189 ms | 202.97.52.25 |
| 16 | 177 ms | 177 ms | 195 ms | 202.97.53.245 |
| 17 | 178 ms | 177 ms | 175 ms | 202.97.53.109 |
| 18 | * | * | * | Request timed out. |
| 19 | 175 ms | 175 ms | 232 ms | 106.120.254.18 |
| 20 | 193 ms | 199 ms | 200 ms | 124.127.161.242 |
| 21 | 200 ms | 201 ms | 201 ms | 124.127.207.2 |

China Networks Internet eXchange Beijing.

Hop 18 does not respond to ICMP packets but allows them to pass on hence the * * *.

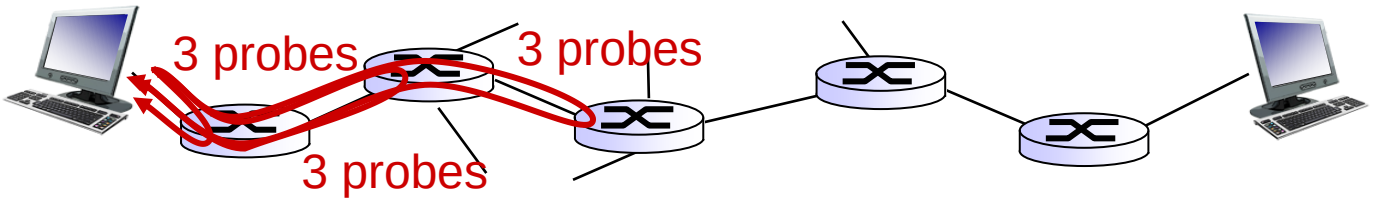

Command Prompt
Microsoft Windows [Version 10.0.18362.1139]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\Users\micha>tracert www.bupt.edu.cn
Tracing route to vn46.bupt.edu.cn [211.68.69.240]
over a maximum of 30 hops:
1 2 ms 1 ms 2 ms dsldevice.lan [192.168.1.254]
2 * * * Request timed out.
3 9 ms 8 ms 10 ms hu0-3-0-4.agr21.lhr01.atlas.cogentco.com [149.6.9.57]
4 10 ms 8 ms 9 ms be3671.ccr51.lhr01.atlas.cogentco.com [130.117.48.137]
5 10 ms 11 ms 9 ms be3487.ccr41.lon13.atlas.cogentco.com [154.54.60.5]
6 90 ms 87 ms 88 ms be12497.ccr41.par01.atlas.cogentco.com [154.54.56.130]
7 109 ms 89 ms 88 ms be3627.ccr41.jfk02.atlas.cogentco.com [66.28.4.197]
8 87 ms 87 ms 88 ms be2806.ccr41.dca01.atlas.cogentco.com [154.54.40.106]
9 99 ms 102 ms 98 ms be2112.ccr41.atl01.atlas.cogentco.com [154.54.7.158]
10 113 ms 121 ms 112 ms be2687.ccr41.iah01.atlas.cogentco.com [154.54.28.70]
11 128 ms 128 ms 127 ms be2927.ccr21.elp01.atlas.cogentco.com [154.54.29.222]
12 136 ms 136 ms 136 ms be2929.ccr31.phx01.atlas.cogentco.com [154.54.42.65]
13 148 ms 148 ms 154 ms be2931.ccr41.lax01.atlas.cogentco.com [154.54.44.86]
14 148 ms 148 ms 148 ms be3271.ccr41.lax04.atlas.cogentco.com [154.54.42.102]
15 179 ms 148 ms 158 ms 38.88.196.186
16 304 ms 350 ms 350 ms 101.4.117.169
17 317 ms 317 ms 350 ms 101.4.117.97
18 307 ms 332 ms 347 ms 101.4.116.81
19 334 ms 348 ms 323 ms 101.4.113.109
20 298 ms 333 ms 348 ms 101.4.112.98
21 349 ms 389 ms 309 ms 101.4.113.65
22 350 ms 343 ms 313 ms 202.112.42.2
23 348 ms 334 ms 304 ms 69.68.211.in-addr.arpa.69.68.211.in-addr.arpa [211.68.69.240]
Trace complete.
C:\Users\micha>

# 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 nth 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



# Test your understanding

With regard to Internet Control Message Protocol (ICMP), which of the following is FALSE?

A. Is a mandatory part of an Internet Protocol (IP) implementation
B. Messages are carried over User Datagram Protocol
C. ICMP does not generate error messages in response to ICMP problems
D. Messages are always directed to the source IP address

# Test your understanding

With regard to Internet Control Message Protocol (ICMP), which of the following is FALSE?

A. Is a mandatory part of an Internet Protocol (IP) implementation
B. Messages are carried over User Datagram Protocol
C. ICMP does not generate error messages in response to ICMP problems
D. Messages are always directed to the source IP address

# What have we learned?

- Routing within Autonomous System
  - Open Shortest Path First (link-state routing)
- Routing between Autonomous Systems
  - Border Gateway Protocol
  - iBGP – within the interior of an AS – distributes list of which addresses take which exit
  - eBGP – at outside of AS calculates routes between AS
- Software defined networking control plane
  - Controller communicates to switches
  - Sends rules, receives statistics and query packets
  - Controller can implement many functions – routing, firewall, switching.
- ICMP
  - used by hosts & routers to communicate network-level information