Computer Networks
COL 334/672

IP Addressing and Inter-domain Routing

Slides adapted from KR

Sem 1, 2025-26

Mark your attendance on Acadly

Recap: Intra-domain routing

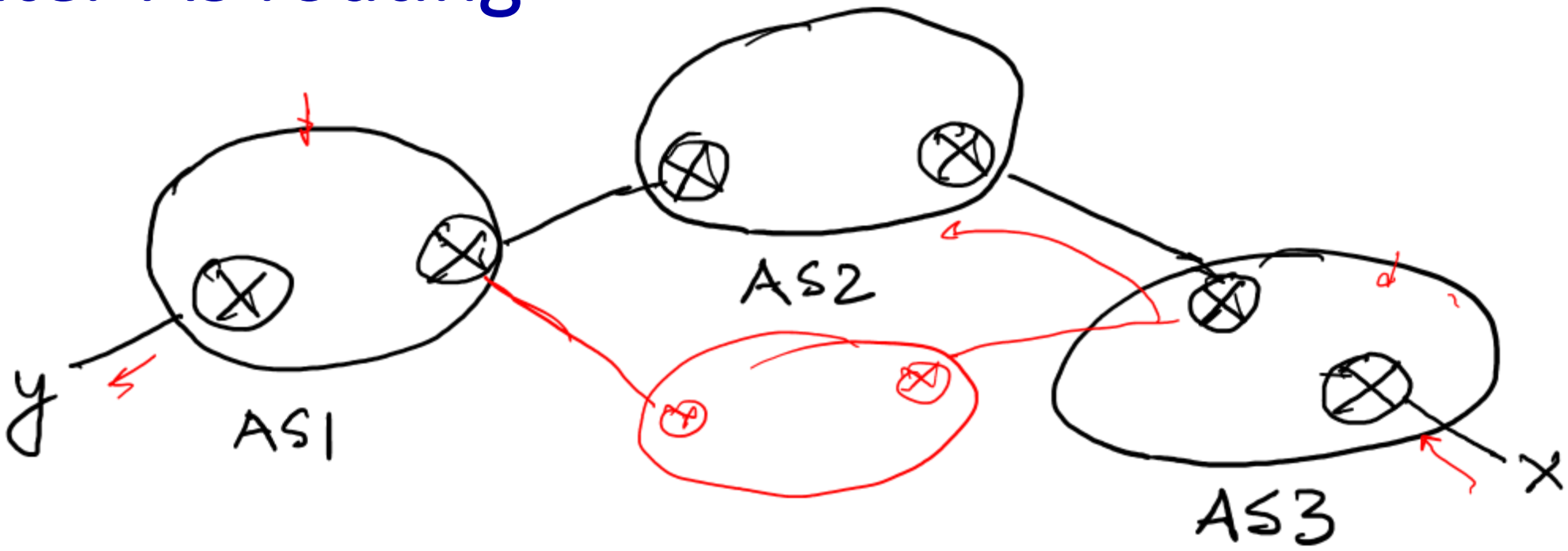
Routing within a network or Autonomous System (AS)

- Finding *shortest (lowest-cost)* paths inside an AS
- Loop-free routing
- Adapt quickly to topology changes

Key approaches

- Distance Vector routing:
  - Routers share distance estimate to destinations with neighbors.
  - E.g. protocols, Routing Information Protocol (RIP), EIGRP: Enhanced Interior Gateway Routing Protocol
- Link State routing:
  - Routers flood topology info, each router uses Dijkstra on full map
  - E.g. protocols, IS-IS protocol, Open Shortest Path First (OSPF)

Inter-AS routing



How does Internet route packet from x to y?

What you need:

- Common addressing scheme IP addressing
- Common routing protocol BGP

Internet addressing or IP address

IPv 4

- IP address: 32-bit identifier associated with each host or router *interface*

① globally unique

- dotted-decimal IP address notation:

② Hierarchical

103.27.10.1 = 01100111 00011011 00001010 00000001

103 27 10 1

- How does an AS obtain IP address?
  - Need a coordinating agency
  - Internet Assigned Numbers Authority (IANA) and Internet Corporation for Assigned Names and Numbers (ICANN)

headquartered in the US

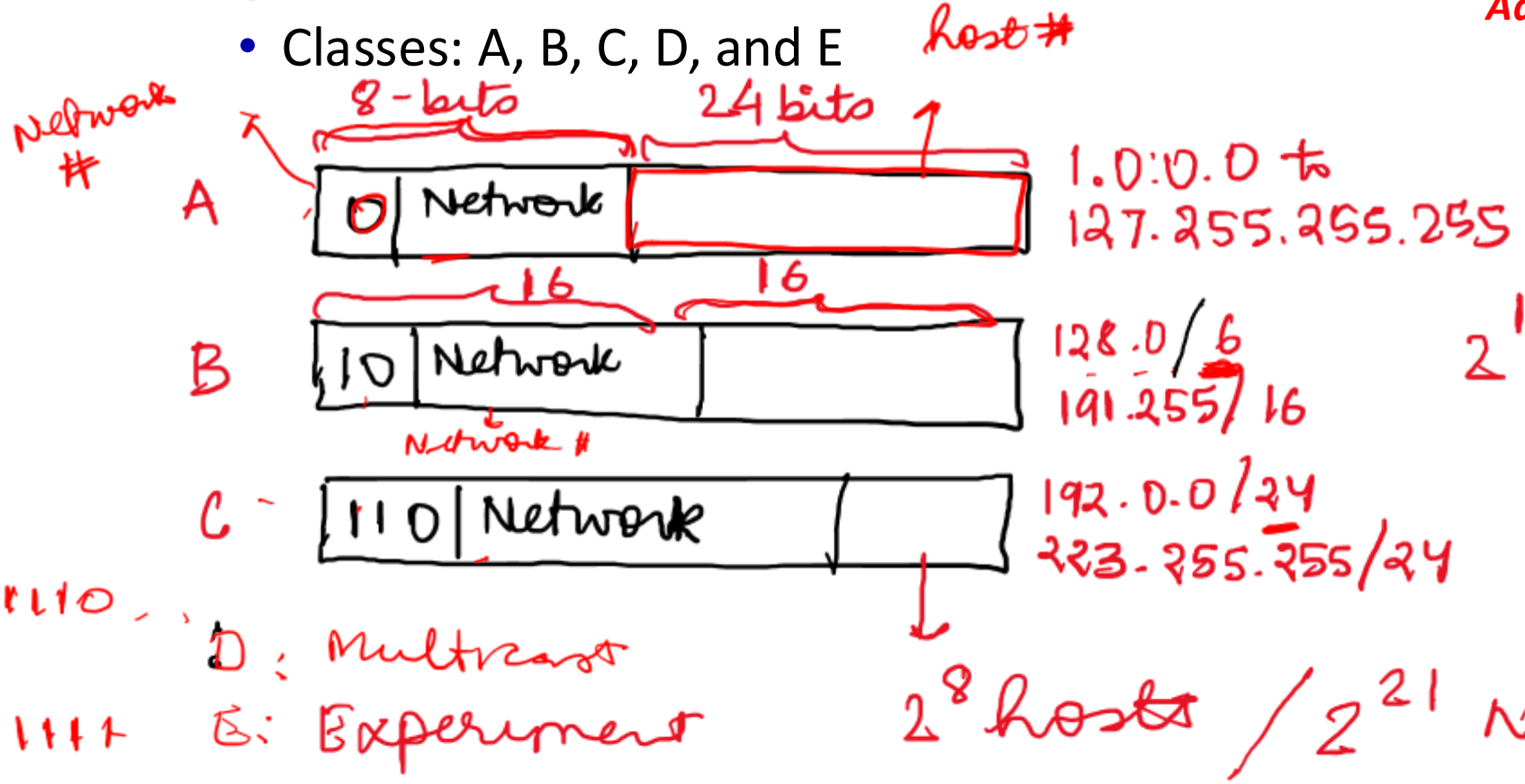
Regional Internet Registeries such as APNIC for Asia

Classful Addressing

Private IP addresses: 10/8, 192.168/16, 172.16/12

- Why do we split the IP address into 4 parts?
- In the beginning, IP addresses were divided into 5 categories, called classful addressing
  - Classes: A, B, C, D, and E

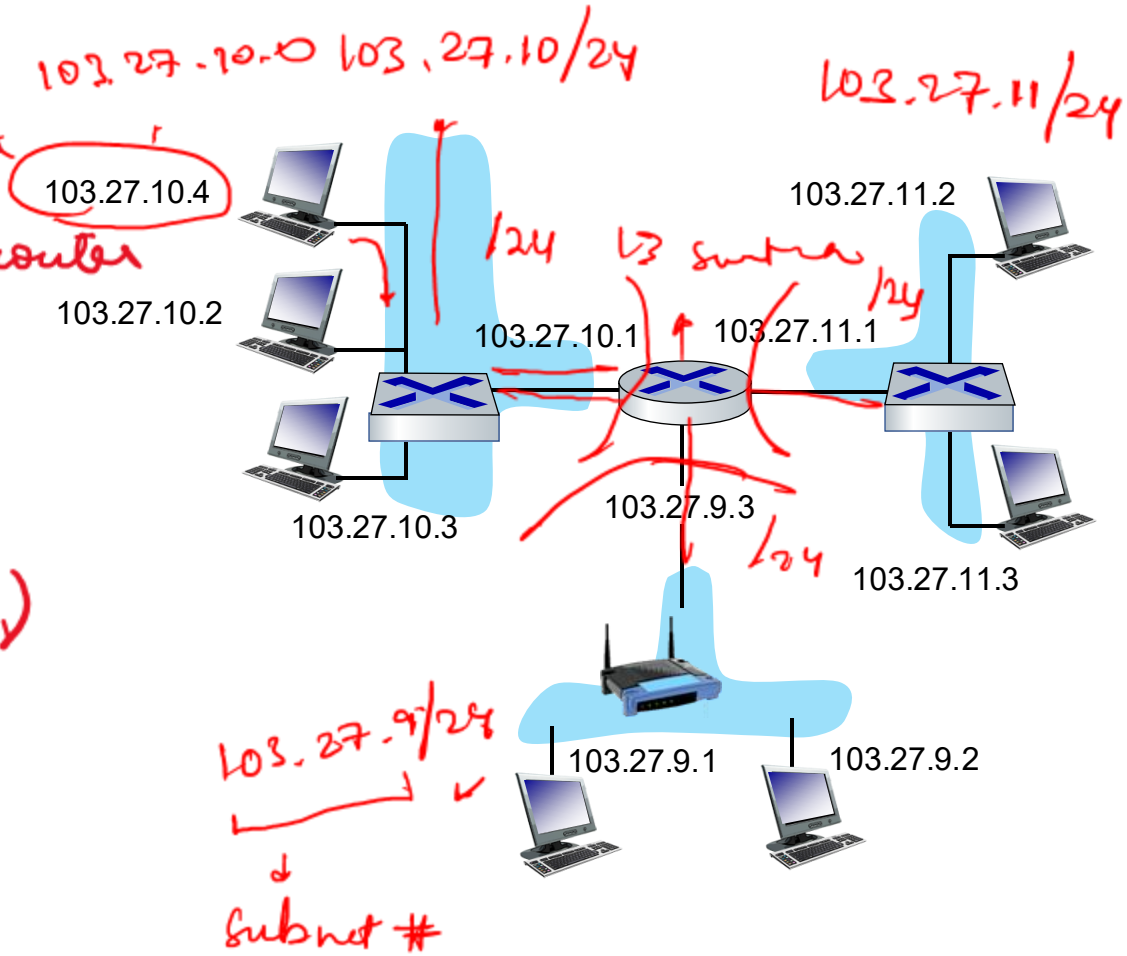
Advantages: **Hierarchical routing**



IP addressing: Subnets

E.g. DST: 103.27.11.1
DST & SUBNET = 103.27.11.0 ≠ 103.27.10.0
Fwd it to gateway router

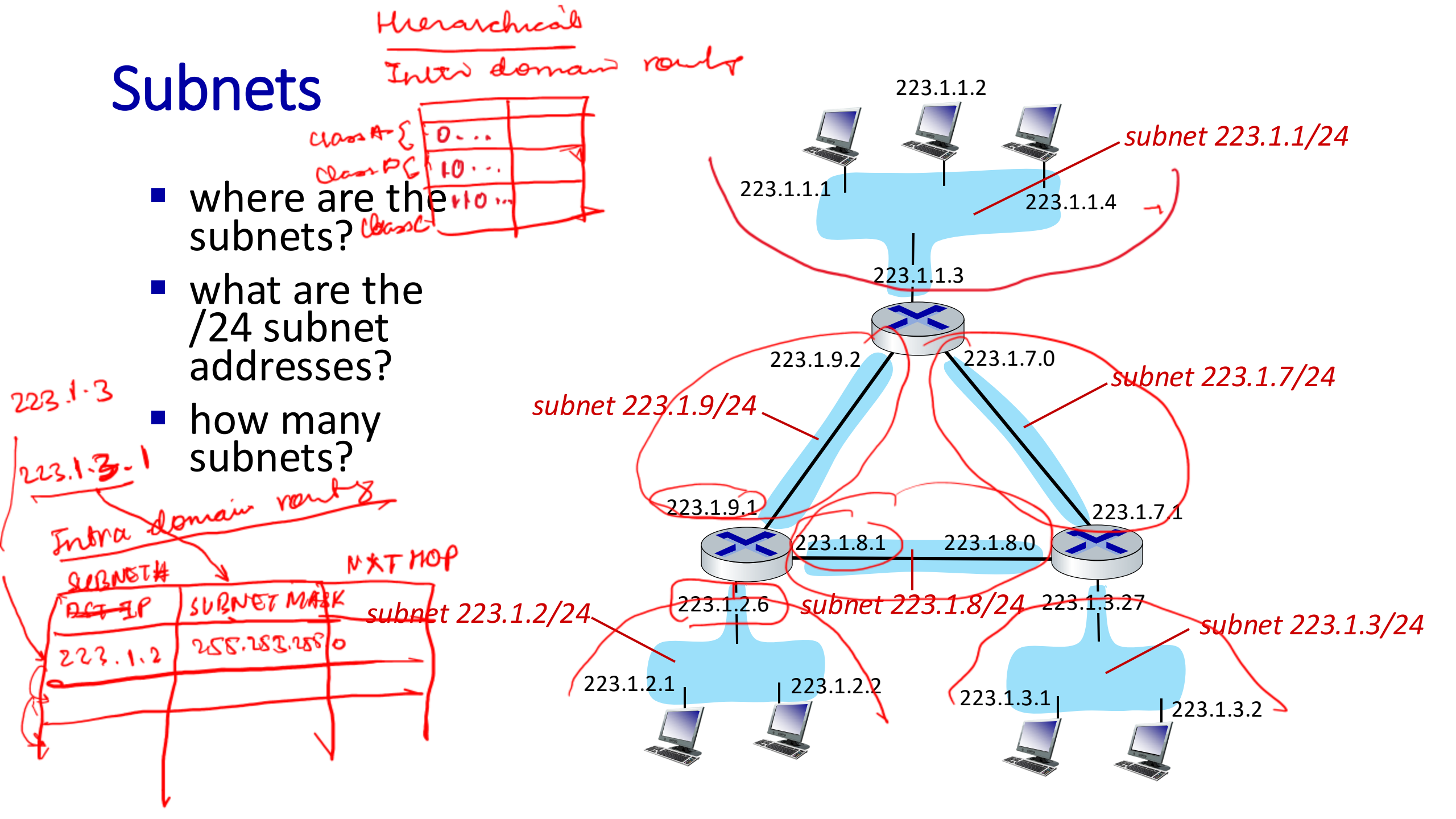
- *What's a subnet ?*
  - device interfaces that can physically reach each other without passing through an intervening router → using L2 routing (e.g. Ethernet)
- IP addresses have structure:
  - subnet part: devices in same subnet have common high order bits
  - host part: remaining low order bits



For DST IP, bitwise & with subnet mask, if subnet # == (DST IP & subnet Mask)
L2 routing · else send it to gateway router

Subnets

- where are the subnets?
- what are the /24 subnet addresses?
- how many subnets?



Limitations of Classful Addressing

- Wastage of IP space

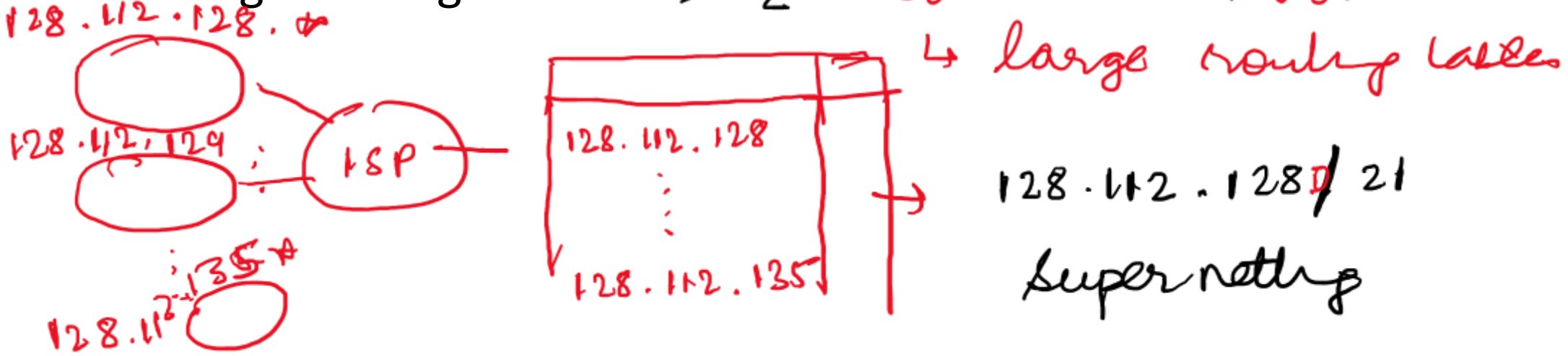
limited options : either 4M, 65000 or 256

- No support of variable subnets

Only /16 or /24 subnets

- Large routing tables → 2^{21} class C Networks

↳ large routing tables



128.112.128.0/21

Supernetting

Example: IP Addresses of two AS

AS3 – Massachusetts Institute of Technology

Country	United States
Website	mit.edu
Hosted domains	718
Number of IPv4	1,836,288
Number of IPv6	6.34×10^{29}
ASN type	Education
Registry	ARIN
Allocated	55 years ago on Jan 01, 1970
Updated	14 years ago on Sep 27, 2010

AS132780 – Indian Institute of Technology Delhi

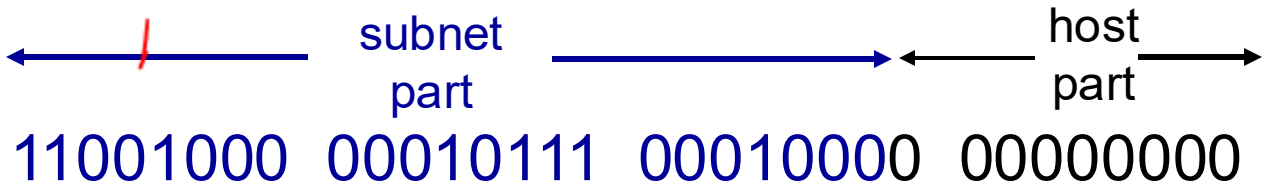
Country	India
Website	iitd.ac.in
Hosted domains	13
Number of IPv4	1,024
Number of IPv6	1.21×10^{24}
ASN type	Education
Registry	APNIC
Updated	2 years ago on Aug 24, 2022

IP addressing: CIDR

CIDR: Classless InterDomain Routing (pronounced "cider")

- subnet portion of address of arbitrary length
- address format: a.b.c.d/x, where x is # bits in subnet portion of address

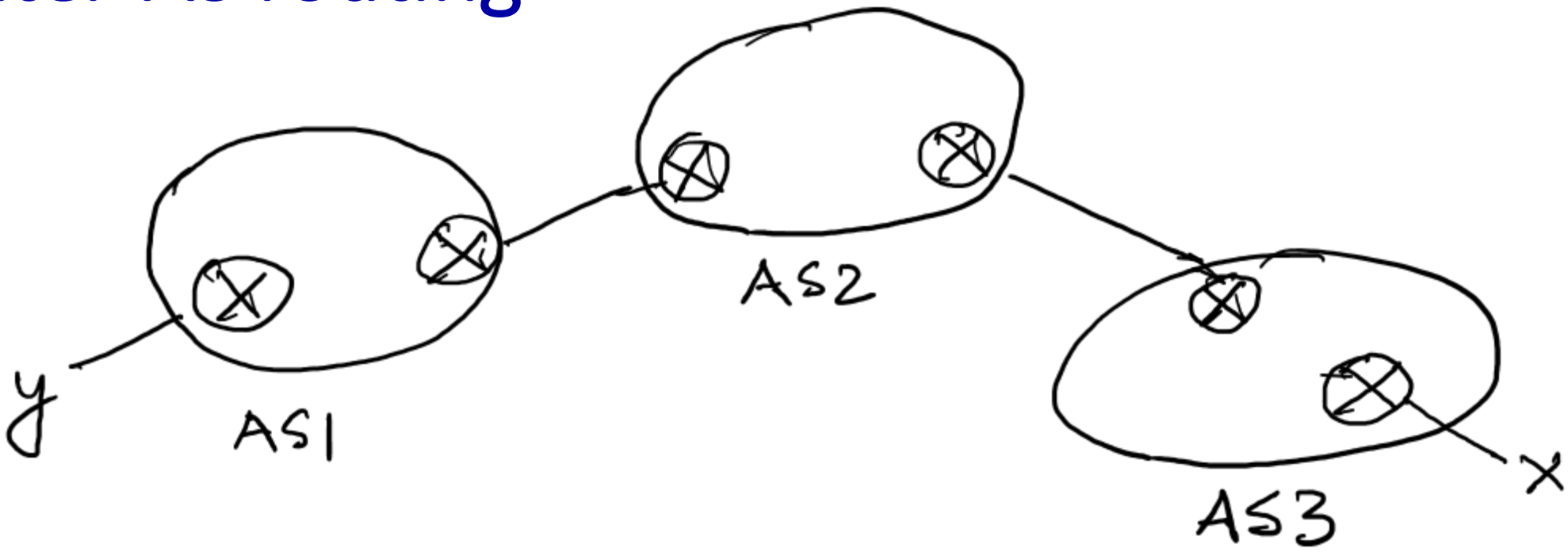
# of bits



200.23.16.0/23

Routers needed to be updated

Inter-AS routing



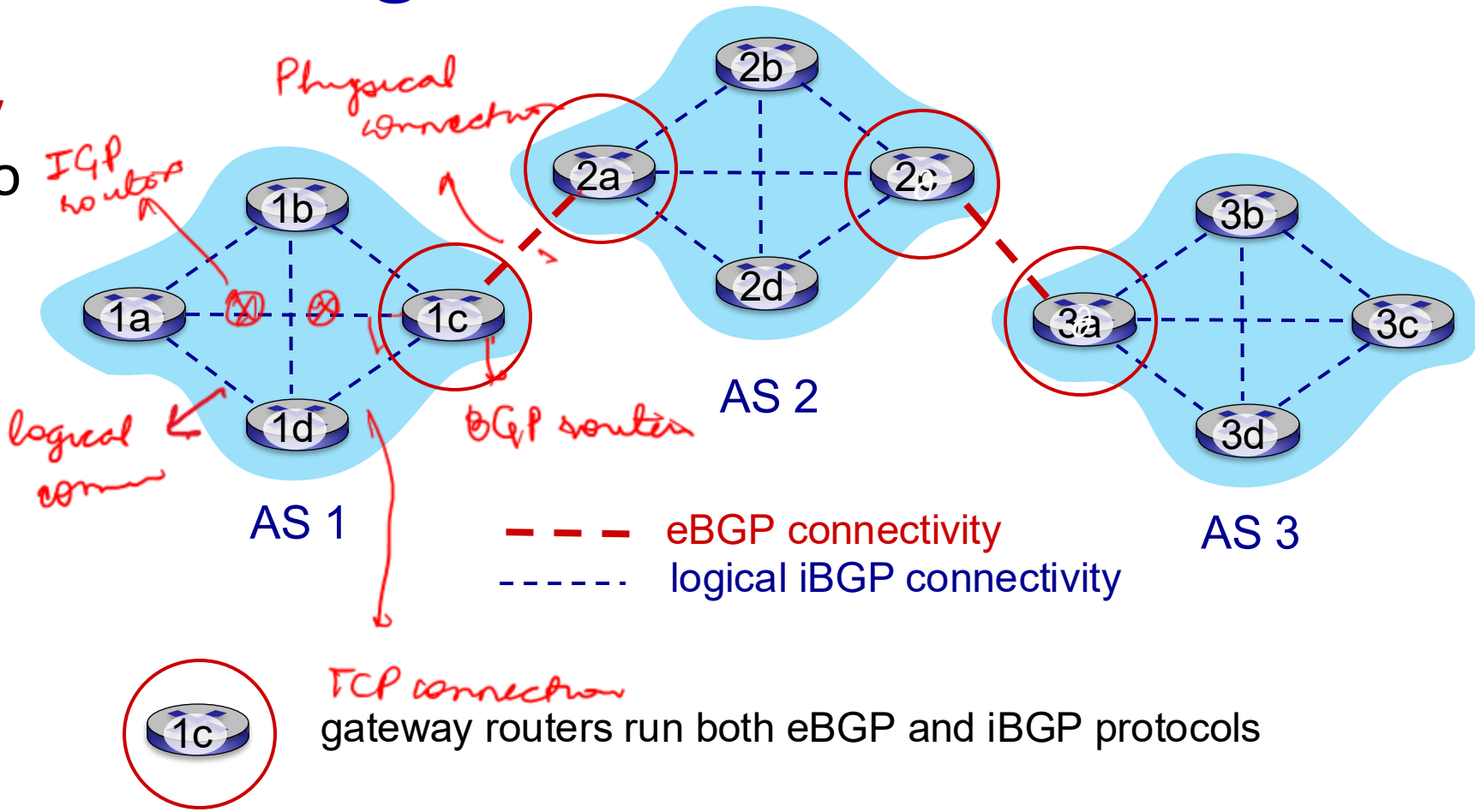
How does Internet route packet from x to y?

What are the requirements for inter-AS routing?

- Common addressing scheme
- **Common routing protocol**

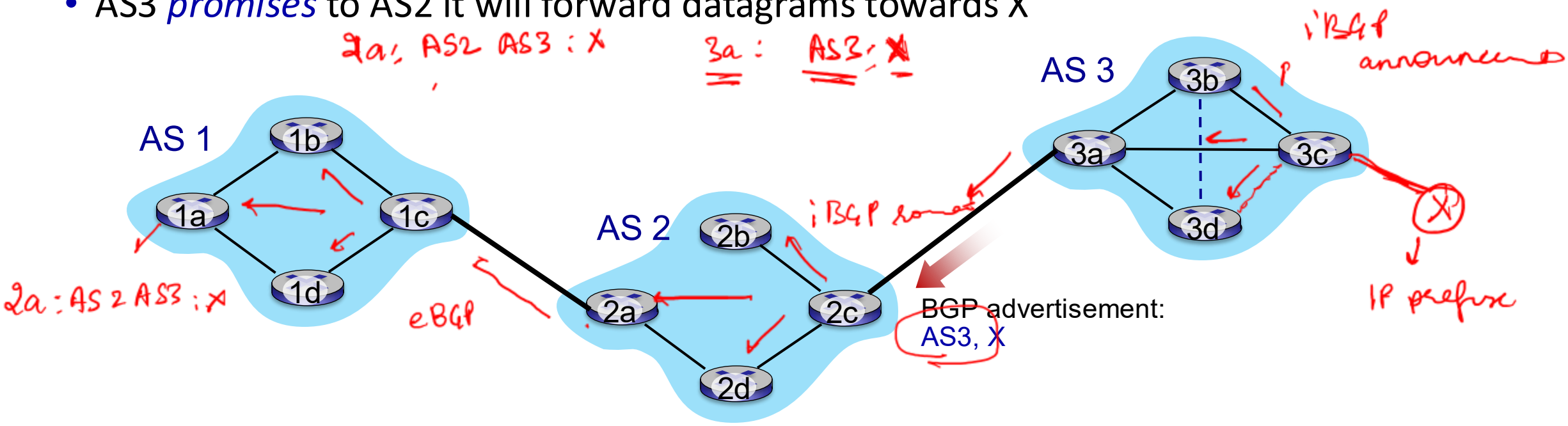
Internet inter-AS routing: BGP

- BGP (Border Gateway Protocol): *the* de facto inter-domain routing protocol
  - "glue that holds the Internet together"
- eBGP, iBGP connections
  - Semi-permanent sessions over TCP
  - Used for exchanging routing information

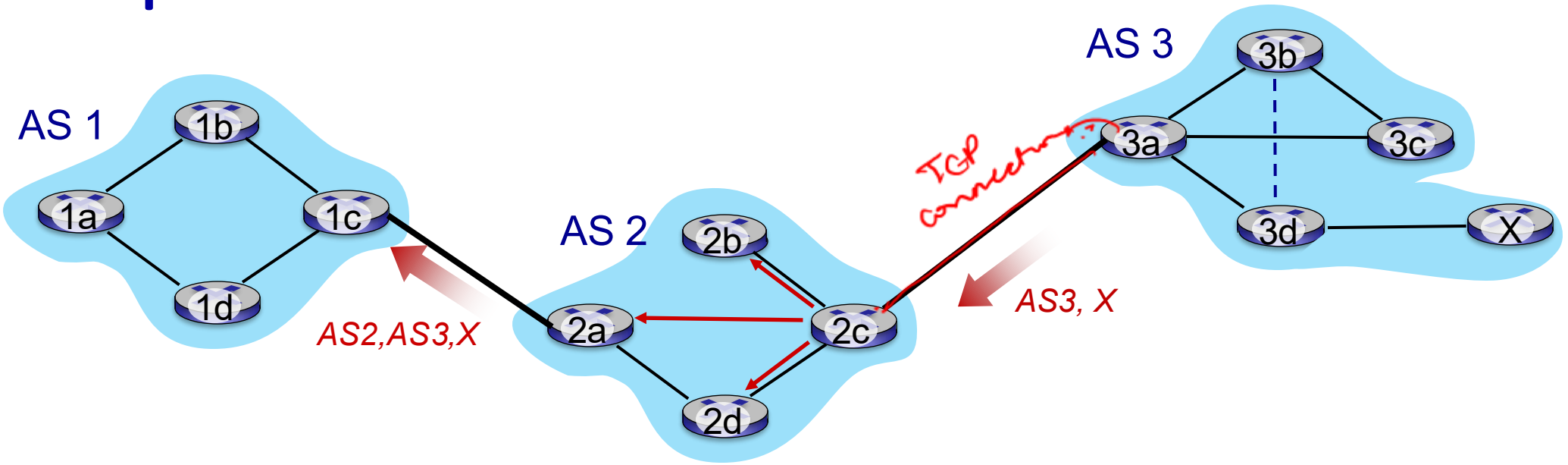


Format of Routing Information

- BGP advertises prefix + attributes
- Two important attributes: AS-PATH and NEXT-HOP
- BGP is a path vector protocol
- When AS3 gateway 3a advertises path AS3,X to AS2 gateway 2c:
  - AS3 *promises* to AS2 it will forward datagrams towards X

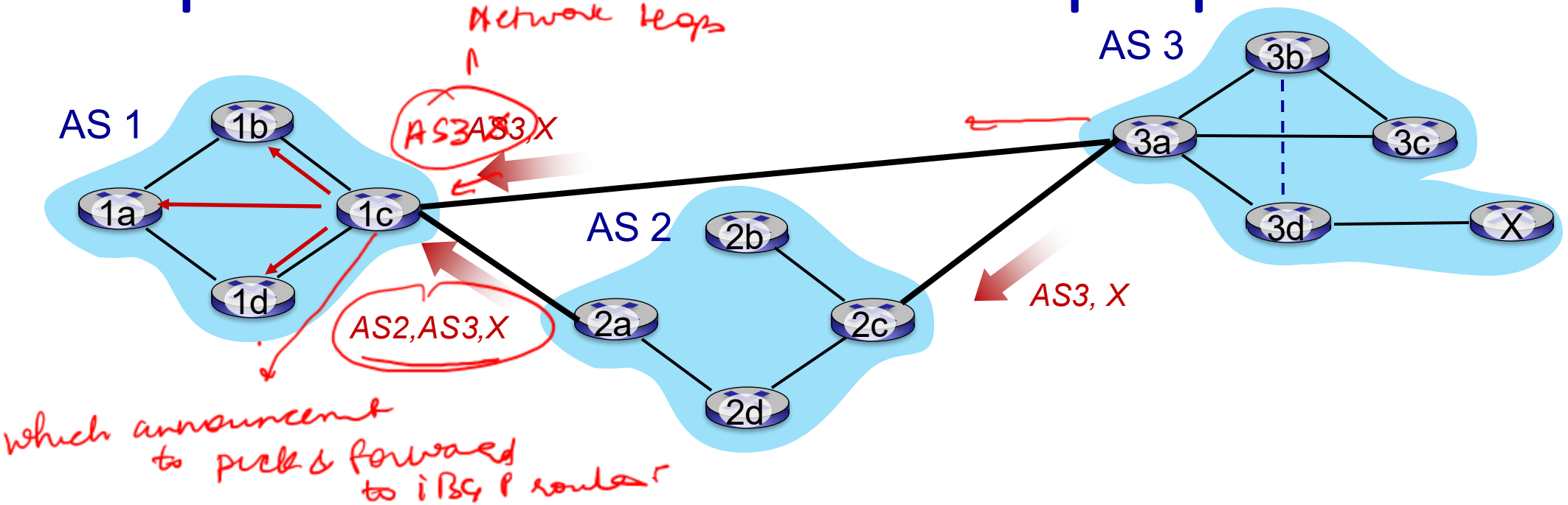


BGP path advertisement



- 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

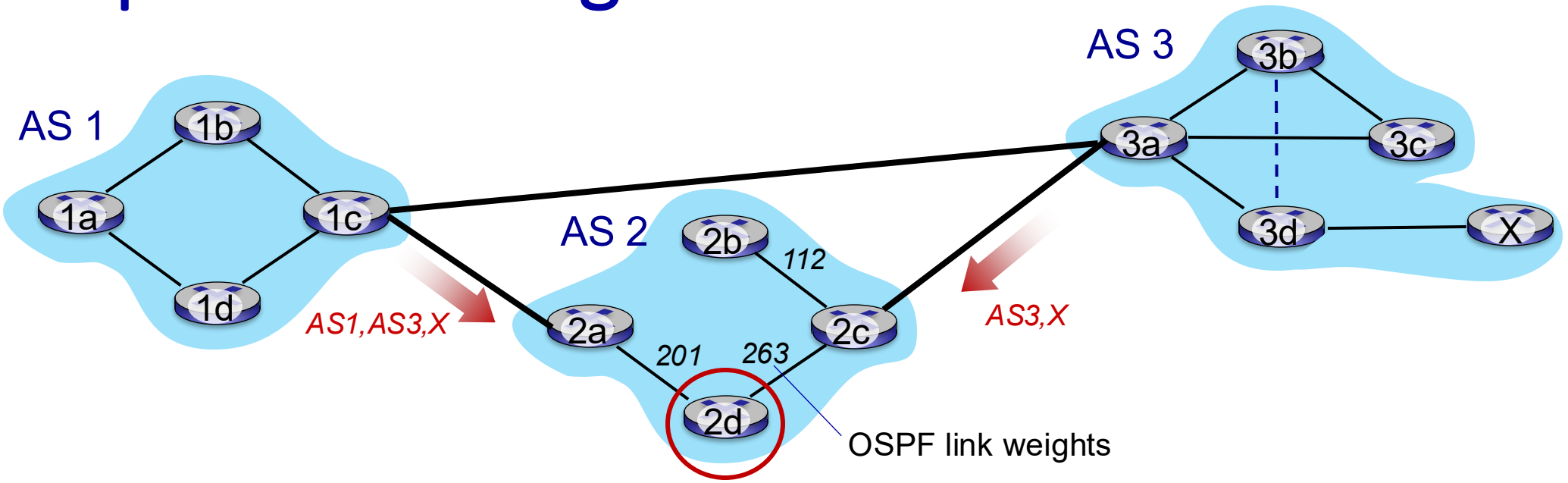
BGP path advertisement: multiple paths



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
- How does AS1 decide which is the best path?

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!

Inter-AS Routing: Policies

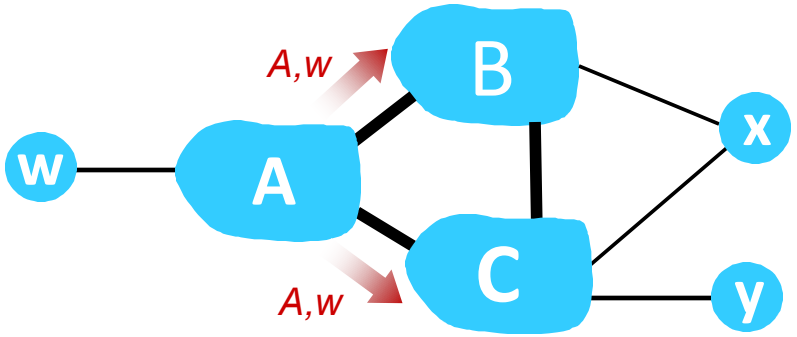
- *No transit traffic through certain ASes*
- *Never put China on a route starting from Ministry of Defence*
- *Do not use Singapore to get from India to Maldives*
- *Traffic starting or ending at Google should not transit through Microsoft*
- *Don’t share information about this peering link to other routers*

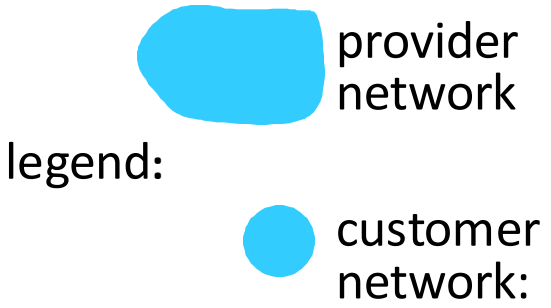
Set local preference on inbound routes

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

BGP: achieving policy via advertisements





ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A advertises path Aw to B and to C
- B *chooses not to advertise* BAw to C!
  - B gets no "revenue" for routing CBAw, since none of C, A, w are B's customers
  - C does *not* learn about CBAw path
- C will route CAw (not using B) to get to w

Why different Intra-, Inter-AS routing ?

policy:

- inter-AS: admin wants control over how its traffic routed, who routes through its network
- intra-AS: single admin, so policy less of an issue

scale:

- hierarchical routing saves table size, reduced update traffic

performance:

- intra-AS: can focus on performance
- inter-AS: policy dominates over performance

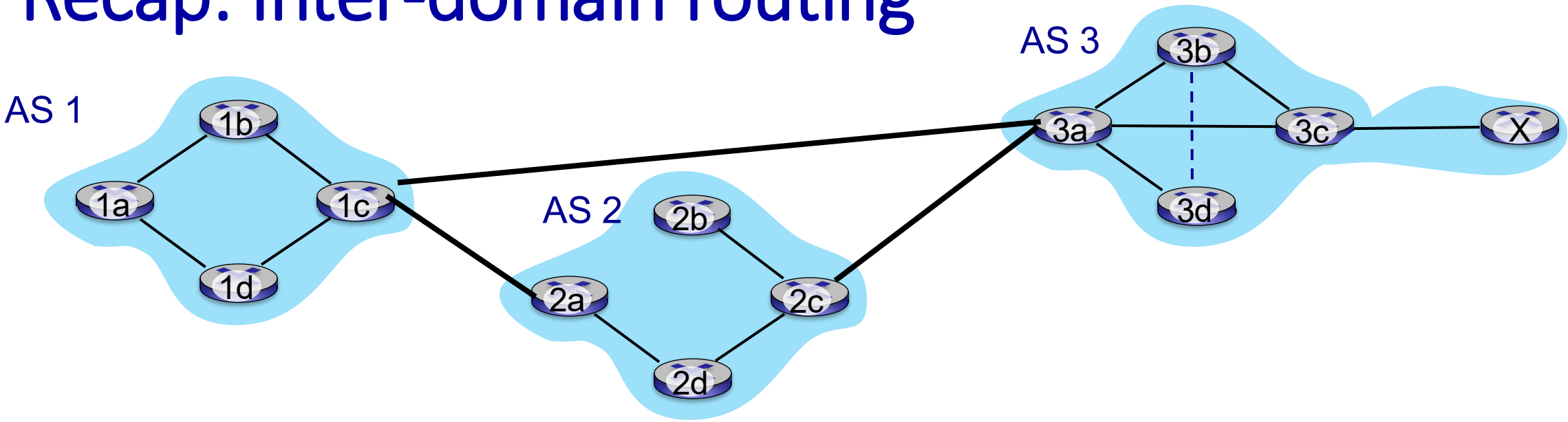
Summary

- Routing Algorithms: Finding “shortest” path from sending host to receiving host
- Intra-domain routing and inter-domain routing
- Intra-domain routing
  - Distance vector (e.g., RIP, EIGRF)
  - Link state (e.g., OSPF)
- Inter-domain routing
  - Path vector routing
  - Border Gateway Protocol (BGP)
- All examples of per-router control plane or a distributed control plane

Recap: Intra-domain Routing

- Algorithms to find shortest path routing within a network
  - Distance vector routing

Recap: Inter-domain routing

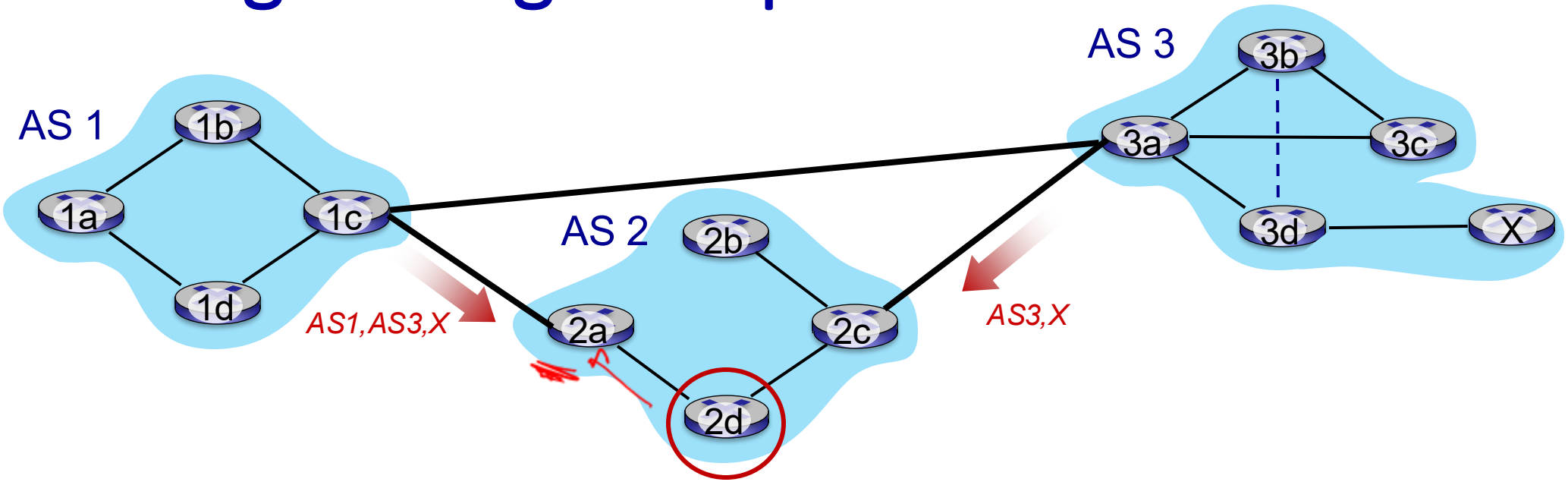


Border Gateway Router (BGP) is the de facto inter-domain routing protocol

- Consists of eBGP and iBGP connections [TCP Connections]
- ASes announce BGP advertisements to neighboring AS
- The advertisement consists of path attributes and IP prefix
- Path attributes include AS path and next hop

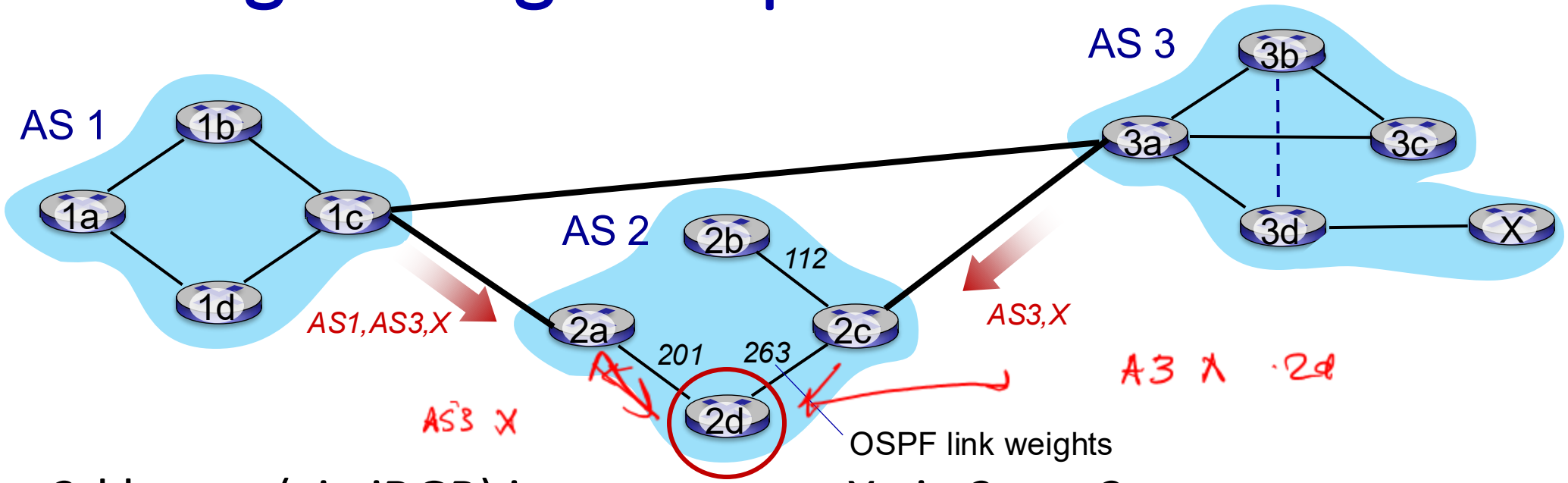
How does a BGP router select among the multiple announcements?

Selecting Among Multiple Announcements



- 2d learns (via iBGP) it can route to X via 2a or 2c
- shortest AS path: choose local gateway that has fewest number of AS hops to the destination

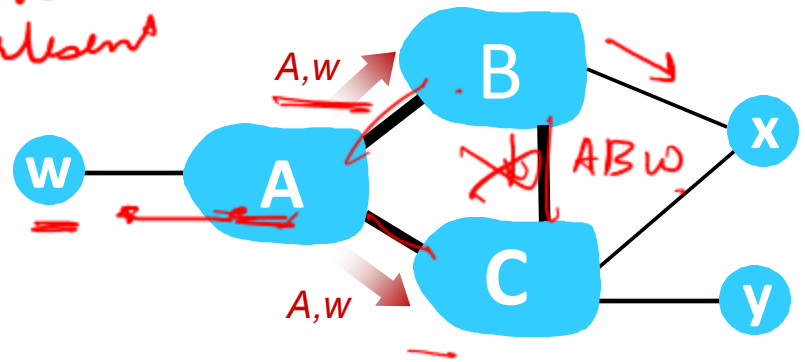
Selecting Among Multiple Announcements

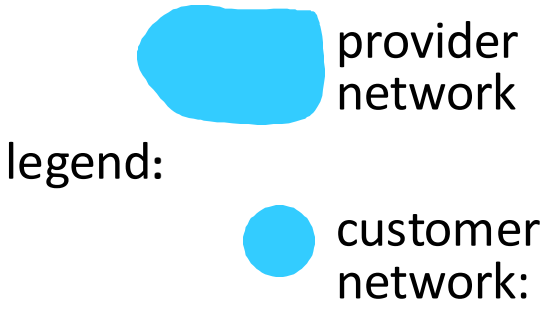


- 2d learns (via iBGP) it can route to X via 2a or 2c
- shortest AS path: choose local gateway that has fewest number of AS hops to the destination
- hot potato routing: choose local gateway that has least *intra-domain* cost (e.g., 2d chooses 2a, even though more AS hops to *X*): don't worry about inter-domain cost!
- network policy: for both path selection and announcements!

BGP: achieving policy via advertisements

Policy
① Route selection
② Route advertisement





ISP only wants to route traffic to/from its customer networks (does not want to carry transit traffic between other ISPs – a typical "real world" policy)

- A advertises path Aw to B and to C
- B *chooses not to advertise* BAw to C!
  - B gets no "revenue" for routing CBAw, since none of C, A, w are B's customers
  - C does *not* learn about CBAw path
- C will route CAw (not using B) to get to w

Inter-AS Routing: Policies

- *No transit traffic through certain ASes*
- *Never put China on a route starting from Ministry of Defence*
- *Do not use Singapore to get from India to Maldives*
- *Traffic starting or ending at Google should not transit through Microsoft*
- *Don't share information about this peering link to other routers*

1. local pref
2. AS path
3. hot potato routing

What is the mechanism in BGP?

Set local preference on inbound routes

BGP route selection

NOC: Network Operation Center

AS2 ↓ 200 AS3 ↑ 100

- 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 ..

cold potato routing

Summary

- Routing Algorithms: Finding "shortest" path from sending host to receiving host
- Intra-domain routing and inter-domain routing
  - Intra: focus more on performance
  - Inter: focus more on policy
- Intra-domain routing
  - Distance vector (e.g., RIP, EIGRF)
  - Link state (e.g., OSPF)
- Inter-domain routing
  - Path vector routing
  - Border Gateway Protocol (BGP)
- All examples of per-router control plane or a distributed control plane

Centralize control plane → SDN
Software Defined Network

Two key network-layer functions

network-layer functions:

- *forwarding:* move packets from a router's input link to appropriate router output link
- *routing:* determine route taken by packets from source to destination
  - *routing algorithms*

analogy: taking a trip

- *forwarding:* process of getting through single interchange
- *routing:* process of planning trip from source to destination

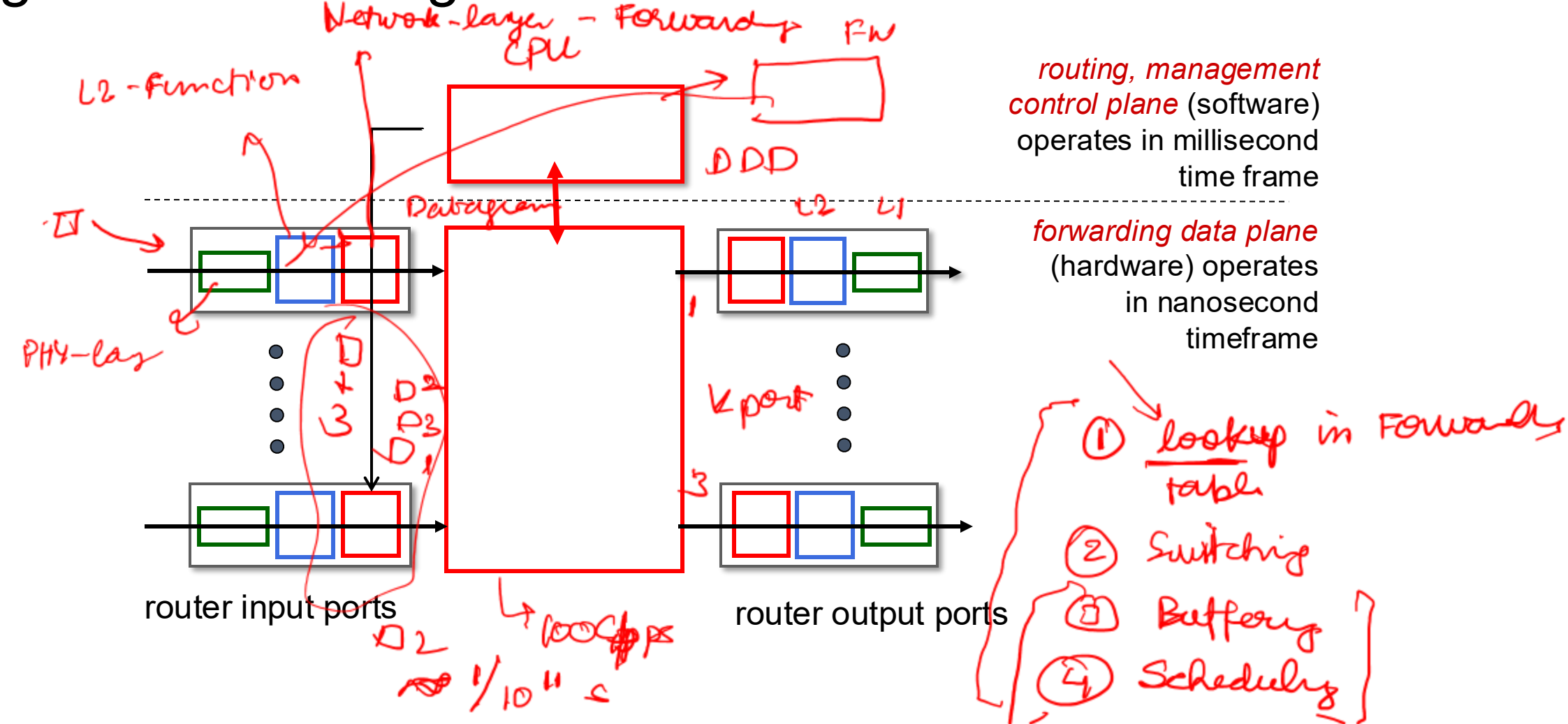
forwarding

routing

SRAM & DRAM

Router architecture overview

high-level view of generic router architecture:



routing, management control plane (software) operates in millisecond time frame

forwarding data plane (hardware) operates in nanosecond timeframe

1. lookup in forwarding table
2. Switching
3. Buffering
4. Scheduling

Lookup

forwarding table

Destination Address Range	Link Interface
11001000 00010111 00010000 00000000 through 11001000 00010111 00010000 00000100	0
11001000 00010111 00010000 00000101 through 11001000 00010111 00010000 00000111	3
11001000 00010111 00011000 11111111	
11001000 00010111 00011001 00000000 through 11001000 00010111 00011111 11111111	2
otherwise	3

Q: but what happens if ranges don't divide up so nicely?

Longest prefix matching

longest prefix match

when looking for forwarding table entry for given destination address, use *longest* address prefix that matches destination address.

Destination Address Range	Link interface
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011*** ********	2
otherwise	3

match!

examples:

11001000 00010111 00010110 10100001 which interface?

11001000 00010111 00011000 10101010 which interface?

Longest prefix matching

longest prefix match

when looking for forwarding table entry for given destination address, use *longest* address prefix that matches destination address.

Destination Address Range	Link interface
11001000 00010111 00010*** ********	0
11001000 00010111 00011000 ********	1
11001000 00010111 00011*** ********	2
otherwise	3

match!

examples:

11001000 00010111 00010110 10100001 which interface?

11001000 00010111 00011000 10101010 which interface?

Longest prefix matching

longest prefix match

when looking for forwarding table entry for given destination address, use *longest* address prefix that matches destination address.

Destination Address Range				Link interface
11001000	00010111	00010***	********	0
11001000	00010111	00011000	********	1
11001000	00010111	00011***	********	2
otherwise				3

match!

examples:

11001000	00010111	00010110	10100001	which interface?
11001000	00010111	00011000	10101010	which interface?