计算机网络

第五章 接口层原理与协议

徐敬东 张建忠

xujd@nankai.edu.cn

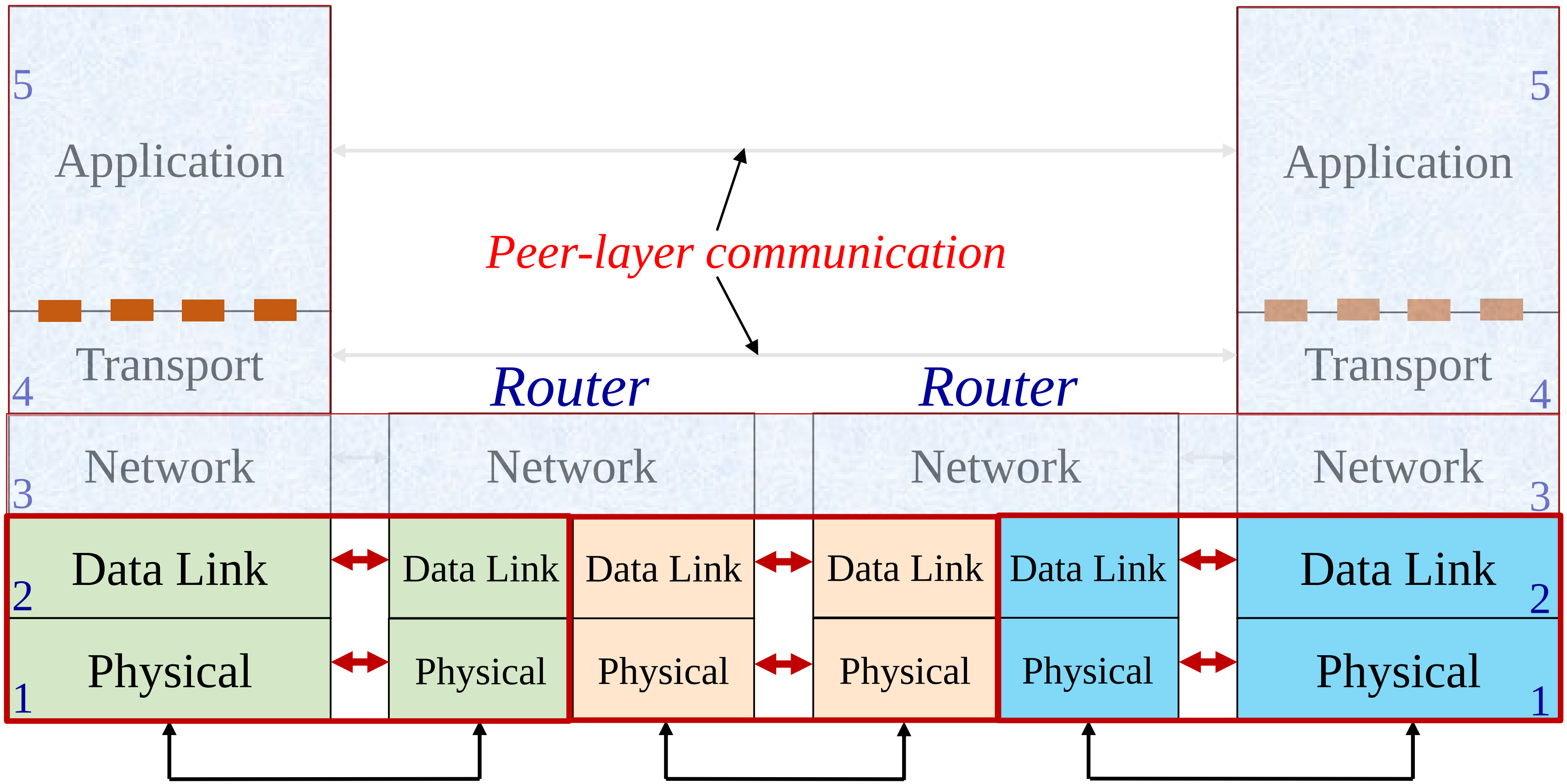
zhangjz@nankai.edu.cn

计算机网络与信息安全研究室

计算机学院&网络空间安全学院

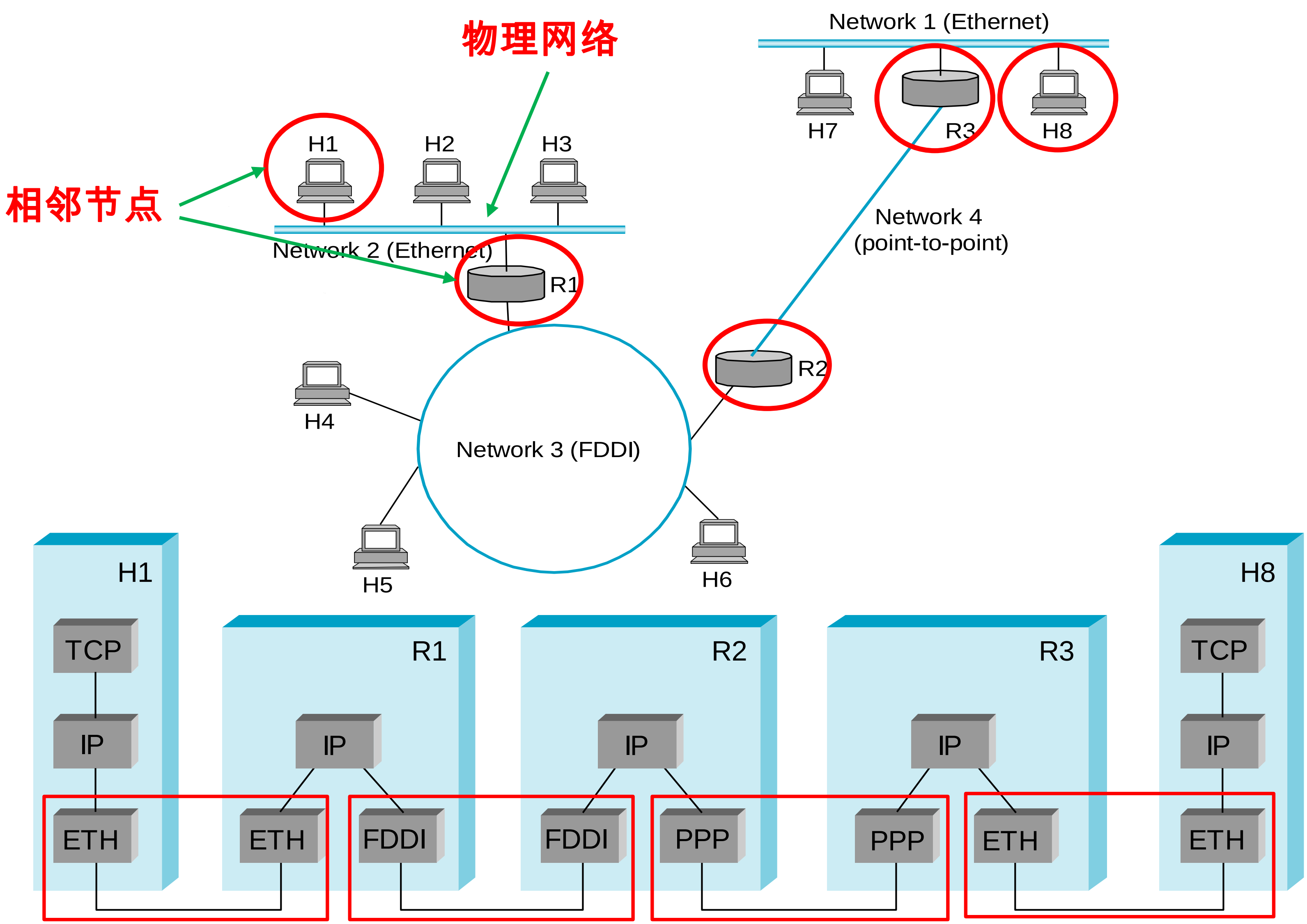
提 纲

Recall: TCP/IP模型

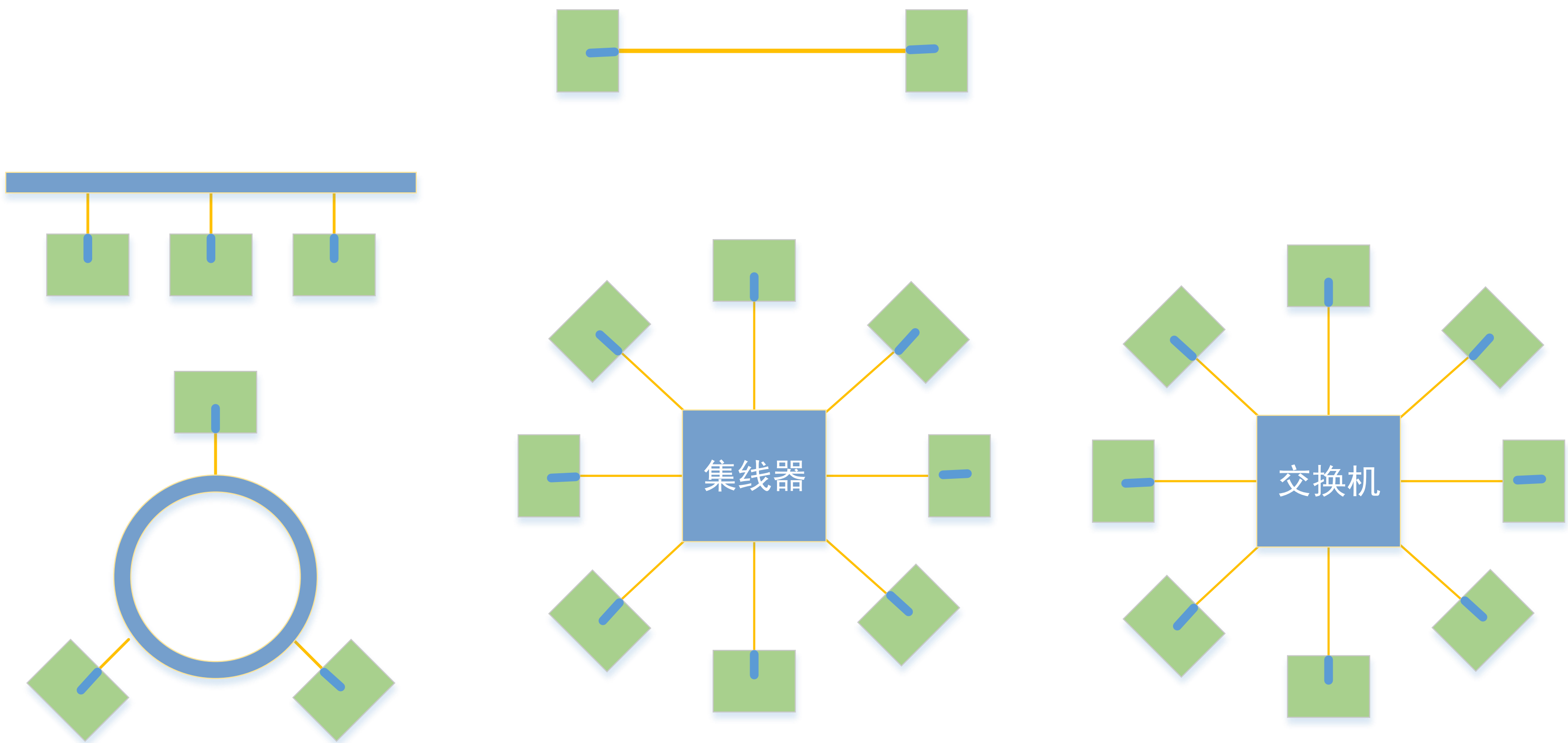


- IP层屏蔽物理网络的差异，向上层提供一致的编址方式和服务，向下支持丰富的接口类型
- 接口层提供同一物理网络中各节点之间的连接和通信

物理网络互联示例


物理网络
相邻节点
Network 1 (Ethernet)
H1
H2
H3
H7
R3
H8
Network 2 (Ethernet)
R1
Network 4
(point-to-point)
R2
H4
Network 3 (FDDI)
H5
H6
H1
TCP
IP
ETH
R1
IP
ETH
FDDI
R2
IP
FDDI
PPP
R3
IP
PPP
ETH
H8
TCP
IP
ETH

物理网络连接方式（拓扑结构）



- 节点到节点连接、共享式连接、交换式连接

5.1 接口层基础

接口层功能

- 物理层：提供位流服务
  - 编码与解码
  - 时钟同步
  - 信号的发送与接收
  - 传输介质和拓扑定义
- 数据链路层：提供可靠或不可靠的传输服务
  - 数据单元及寻址方式定义
  - 链路层差错检测
  - 链路层的复用和分用
  - 可靠数据传输
  - 共享式连接：提供介质访问控制方法

接口层功能通常由网络接口卡(NIC)和驱动程序共同实现

接口层技术分类

- 有多种接口层技术，传统上大致可以分成三类
  - 局域网技术（LAN, Local Area Network ）
    - 如：以太网（Ethernet）、无线局域网（WiFi）
  - 城域网技术 （MAN, Metropolitan Area Network）
    - 如：FDDI、交换式Ethernet
  - 广域网技术Wide Area Network (WAN)
    - 如：ATM
- 其他接口技术
  - 个人区域网
    - 如：蓝牙技术（Bluetooth）
  - 无线传感网络
    - 如：Zigbee技术

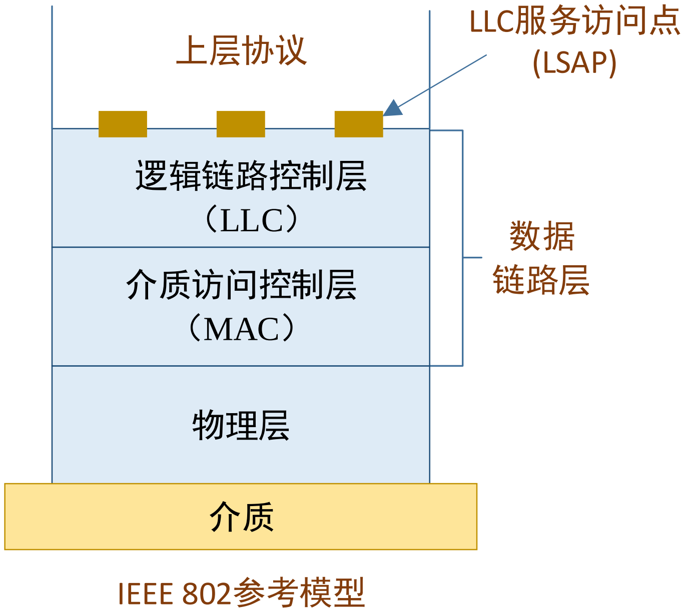
局域网体系结构与数据封装

介质访问控制层 (Medium Access Control)

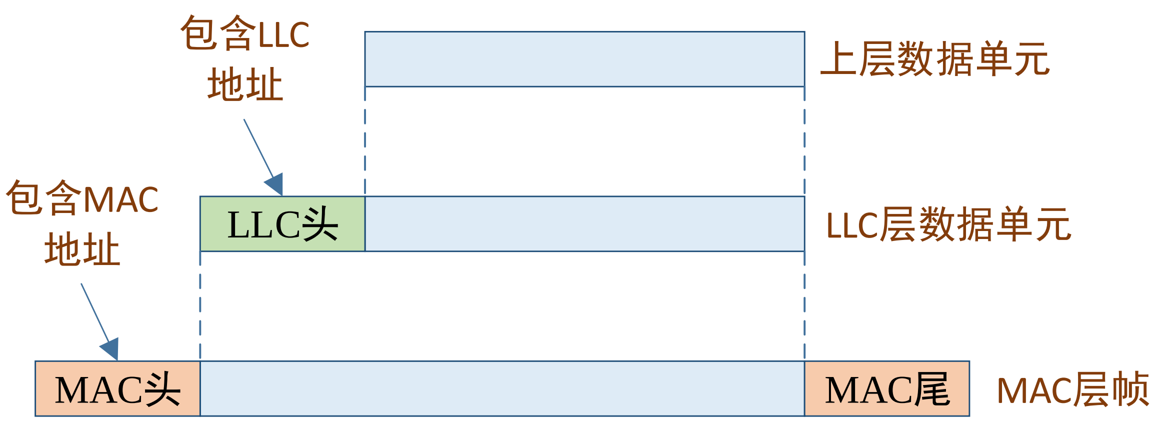
- 物理节点寻址
- 差错控制
- 介质访问控制（共享式连接）

逻辑链路控制层（Logical Link Control）

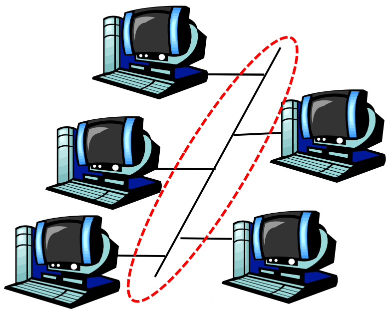
- 链路层的复用和分用
- 可靠数据传输



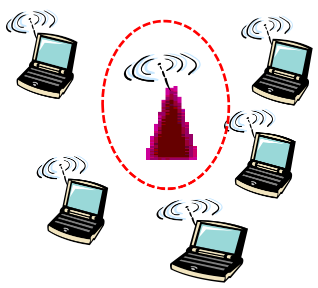
IEEE 802参考模型



- 共享式局域网
  - 网络中任一节点发送的信息会被网络中所有节点收到（广播传输）
    - 例如，共享式以太网，无线局域网，FDDI等
  - 需要协调节点对共享介质的访问
    - **介质访问控制方法**

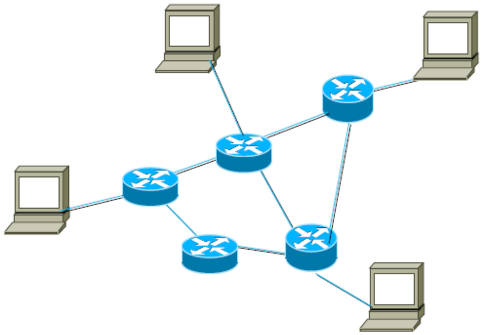


共享式以太网

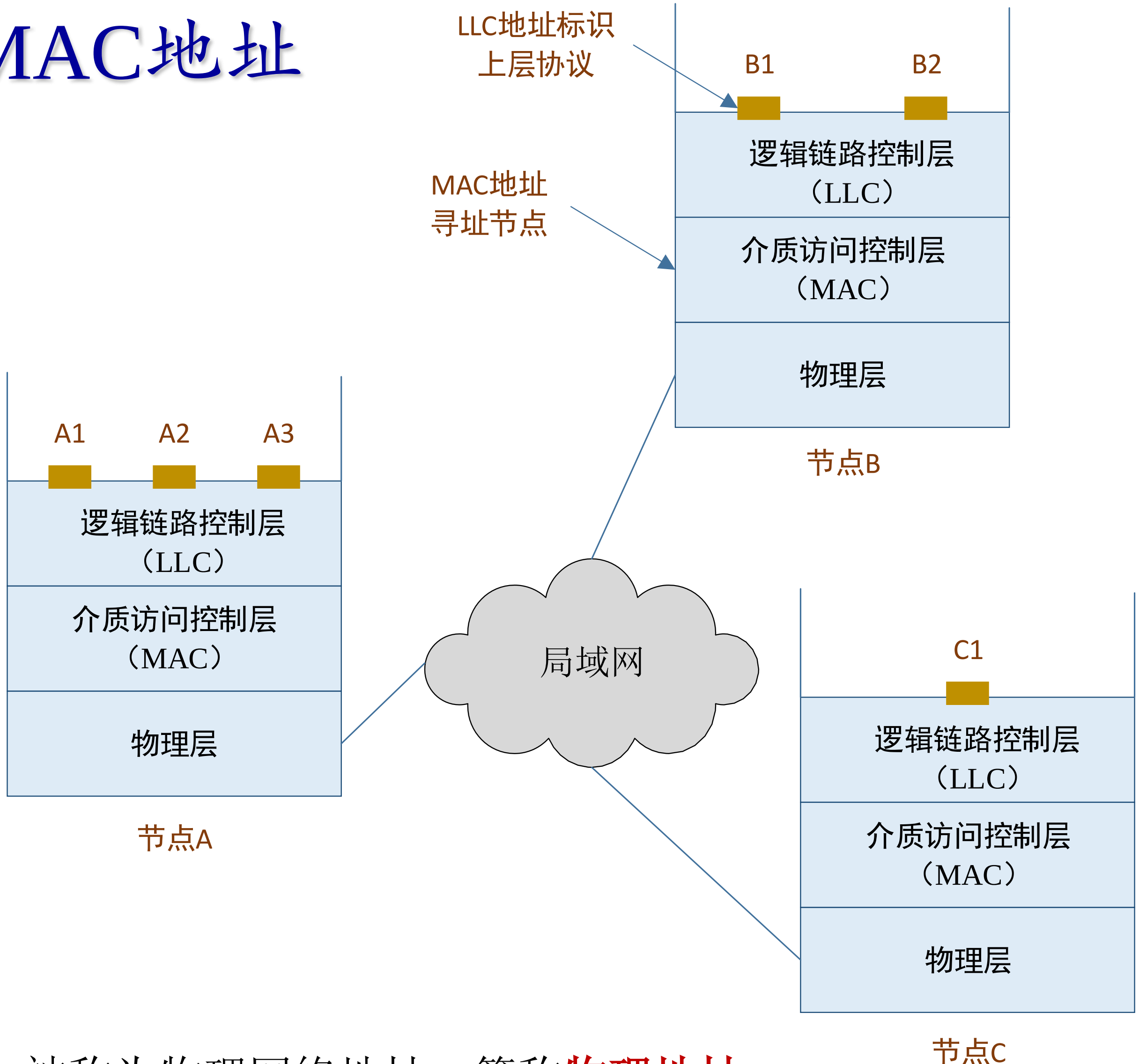


802.11无线局域网

- 交换式局域网
  - 链路采用点到点连接
    - 例如，交换式以太网
  - 交换机成为网络连接的核心
    - **完成链路层数据单元的转发**
    - 通常采用统计多路复用



LLC和MAC地址



- MAC地址：被称为物理网络地址，简称**物理地址**

MAC地址—物理地址

32-bit IP address:

- *network-layer* address
- used to transmit datagram to **destination network** (recall IP network definition)

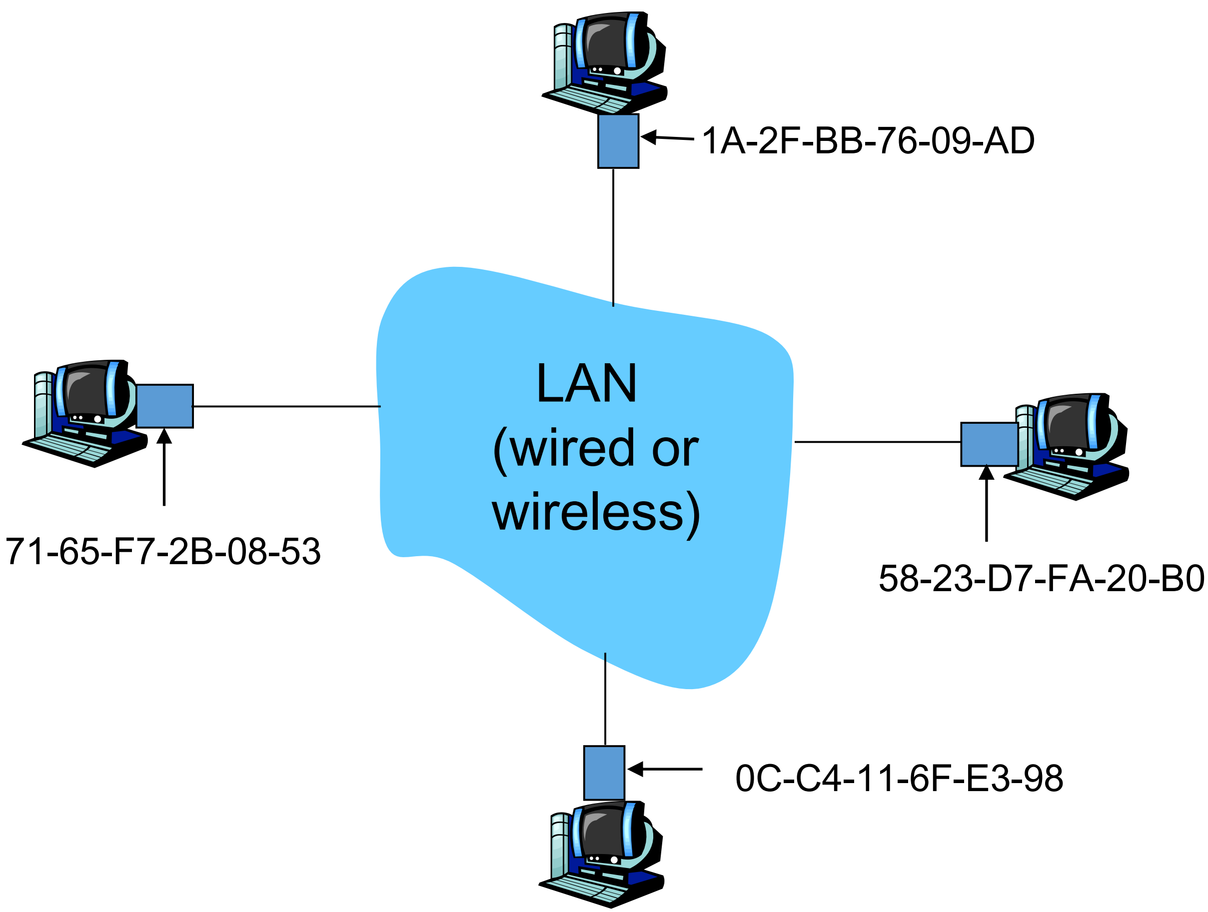
MAC (or physical) address:

- used to transmit datagram from one interface to **another physically-connected interface** (same physical network)
- *48 bit MAC address* (for most LANs)
  - burned in the adapter ROM or EPROM
- 2^{48} possible LAN address (flat address)
  - can move LAN card from one LAN to another

5.3 局域网编址与ARP协议

MAC地址—物理地址

Each NIC on LAN has unique LAN address



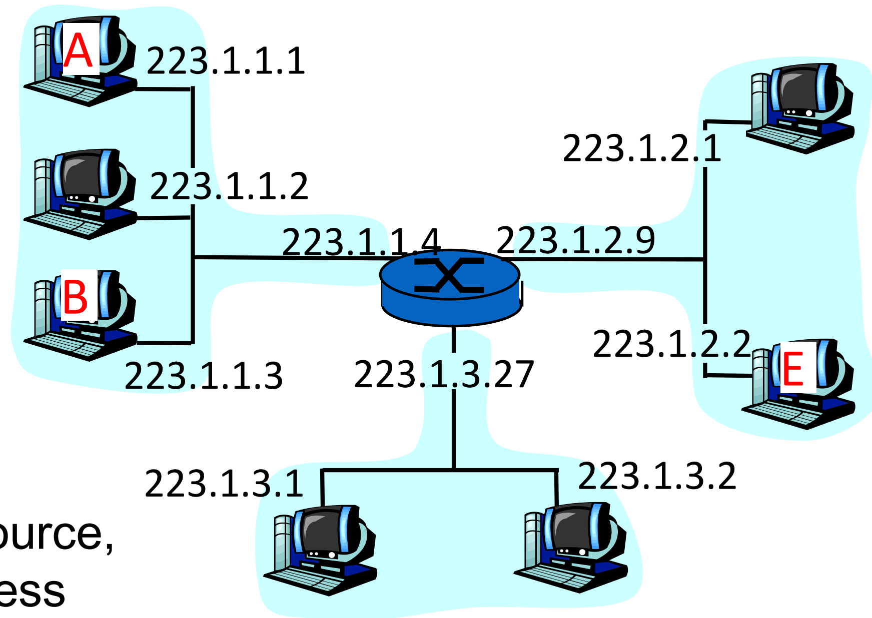
Broadcast address = FF-FF-FF-FF-FF-FF

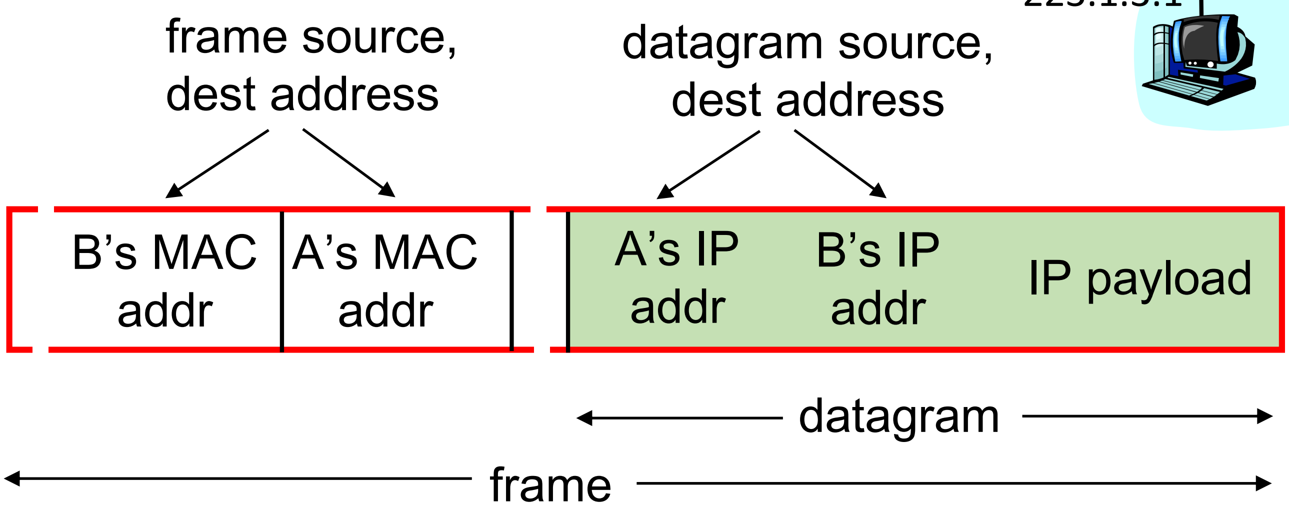
= NIC

Recall earlier routing discussion

Starting at A, given IP datagram addressed to B:

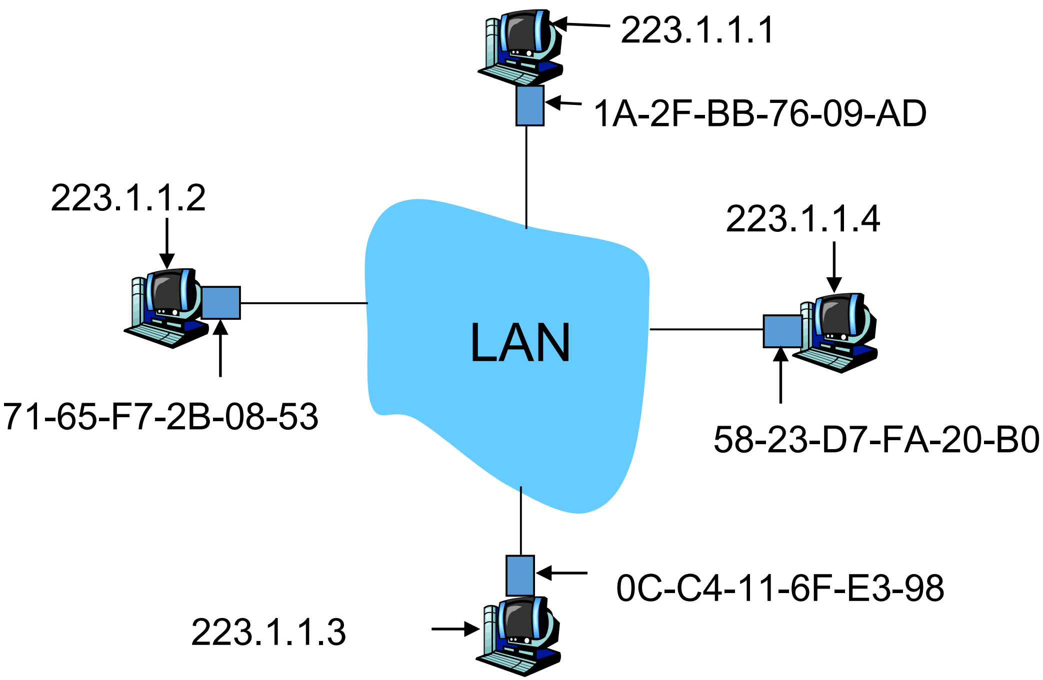
- look up net. address of B, find B on same net. as A
- MAC layer send datagram to B inside MAC-layer frame





ARP: Address Resolution Protocol

Question: how to determine MAC address of B given B's IP address?



- Each IP node (Host, Router) on LAN has ARP table
- ARP Table: IP/MAC address mappings for some LAN nodes

  < IP address; MAC address; TTL>

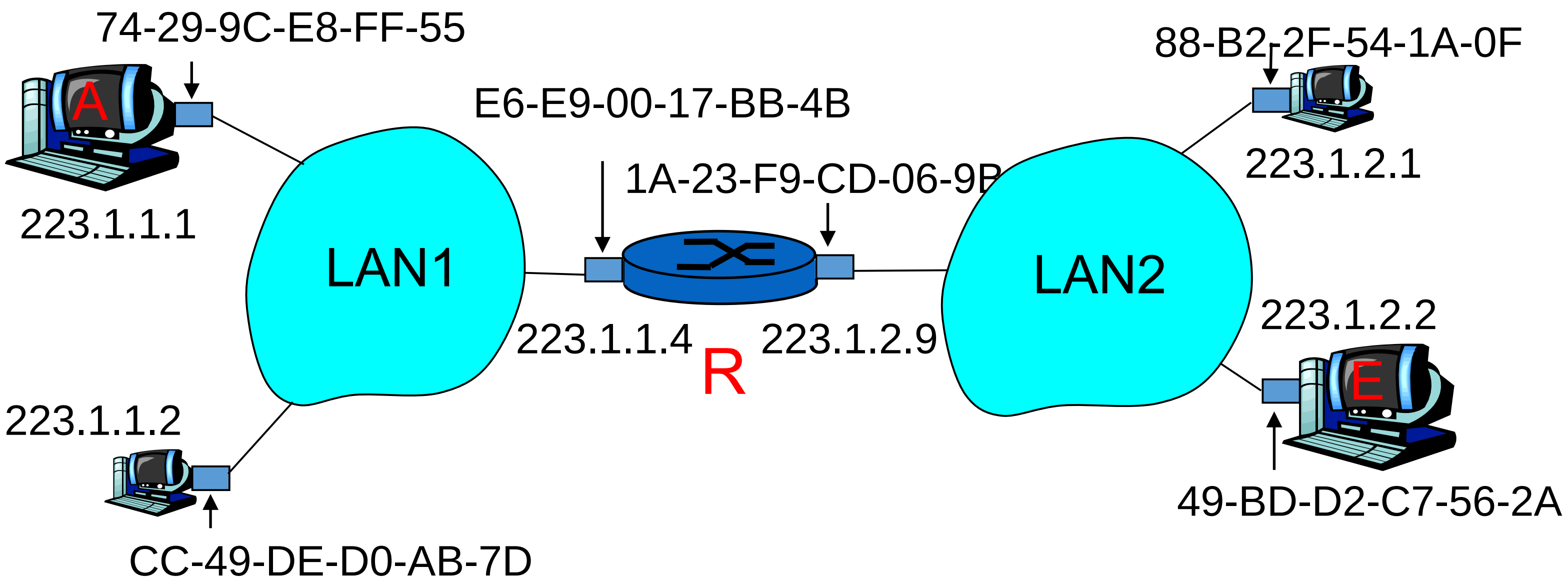
  - TTL (Time To Live): time after which address mapping will be forgotten (typically 20 min)

ARP protocol

- A knows B's IP address, wants to learn physical address of B
- A broadcasts ARP query packet, containing B's IP address
  - all machines on LAN receive ARP query
- B receives ARP packet, replies to A with its (B's) physical address
- A caches (saves) IP-to-physical address pairs in ARP table until information becomes old (timeout)
- ARP is “plug-and-play”:
  - nodes create their ARP tables without intervention from net administrator

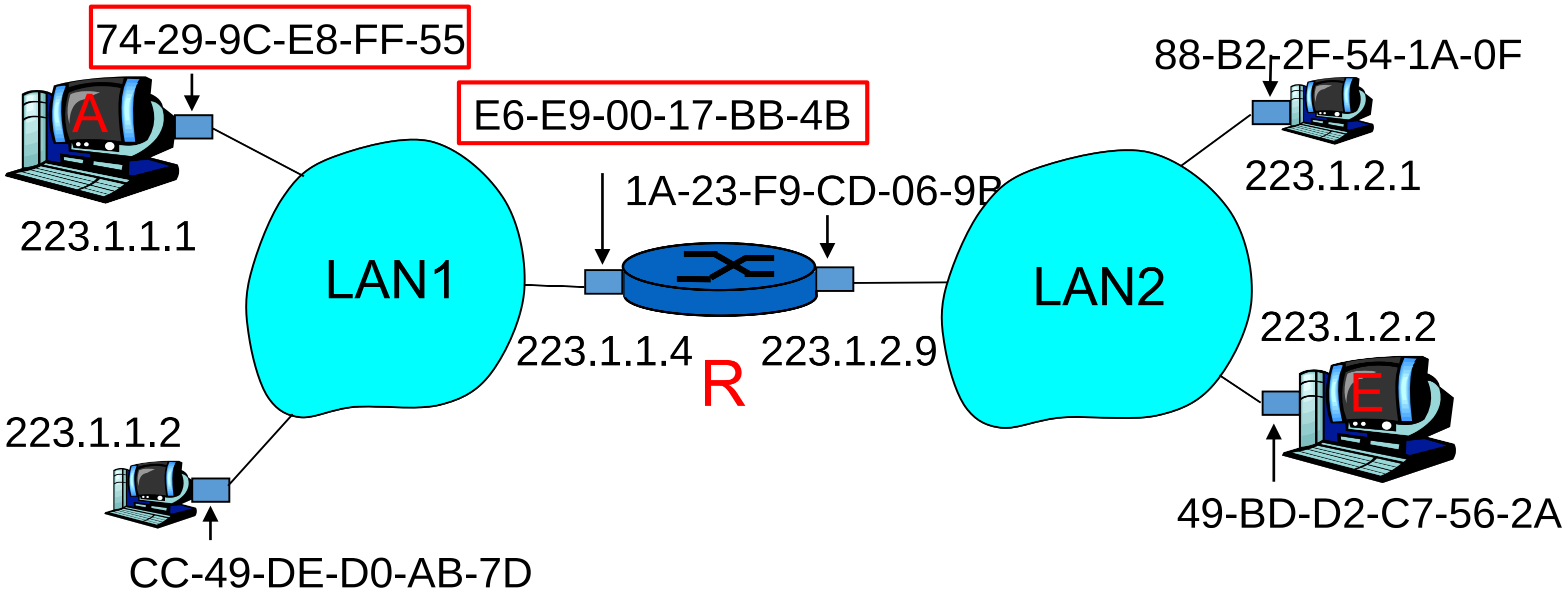
Routing to another LAN

walkthrough: routing from A to E via R



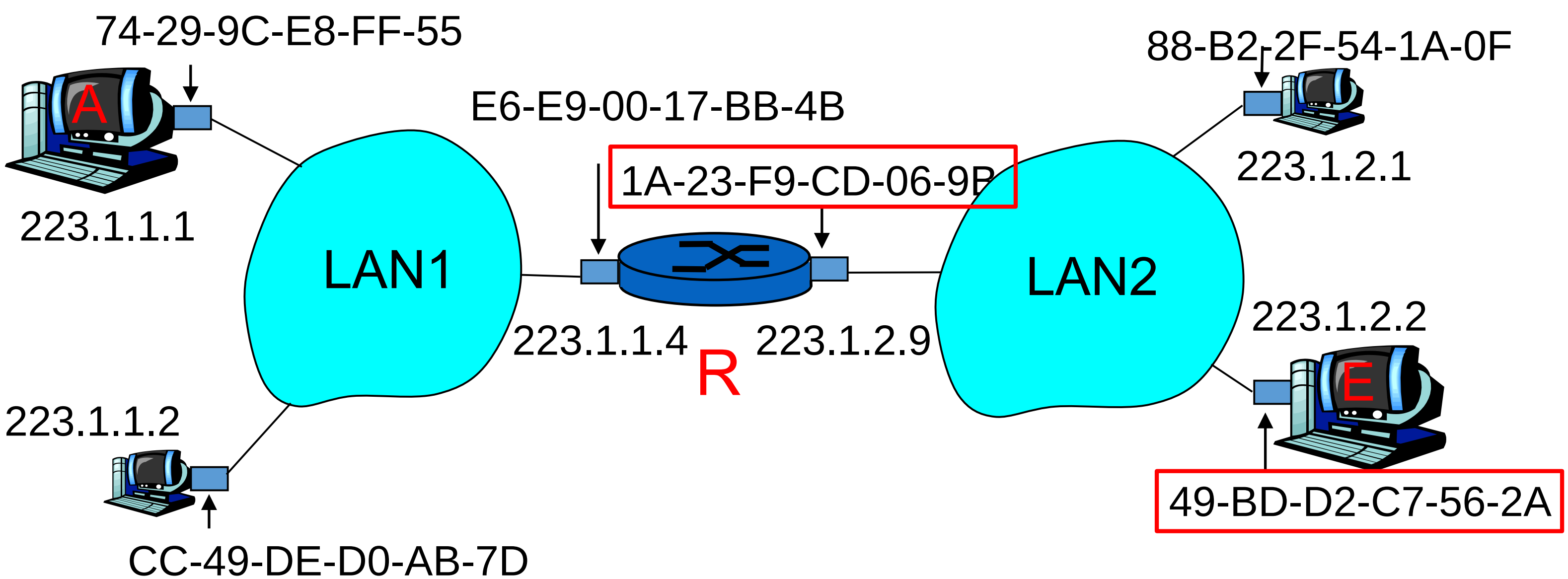
- A creates IP packet with source A, destination E
- In routing table at source Host, find router 221.1.1.4
- A uses ARP to get R's MAC address for 223.1.1.4

Routing to another LAN (cont.)



- A creates Ethernet frame with R's MAC as dest, Ethernet frame contains A-to-E IP datagram
- A’s data link layer sends Ethernet frame
- R’s data link layer receives Ethernet frame

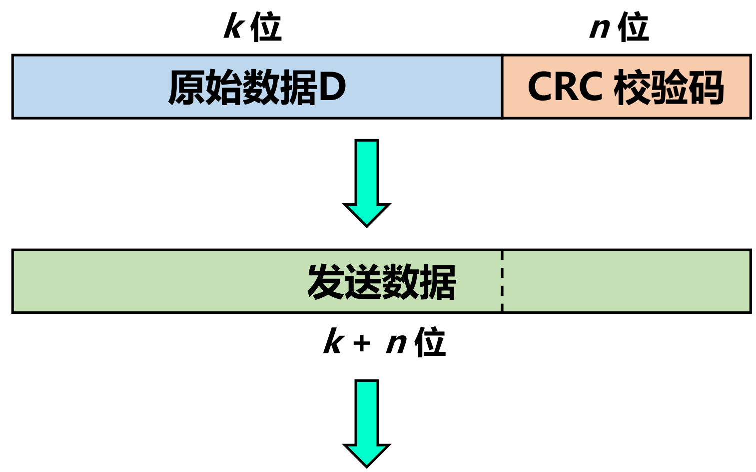
Routing to another LAN (cont.)



- R removes IP datagram from Ethernet frame, sees its destined to E
- R uses ARP to get E's physical layer address
- R creates frame containing A-to-E IP datagram sends to E

循环冗余校验（CRC）

- 在数据链路层中，广泛使用循环冗余校验（Cyclic Redundancy Check ）
- CRC校验码计算方法
  - 设原始数据D为k位二进制位模式
  - 如果要产生n位CRC校验码，事先选定一个n+1位二进制位模式G（称为生成多项式），G的最高位为1
  - 将原始数据D乘以2^n（相当于在D后面添加 n 个 0），产生k+n位二进制位模式，用G对该位模式做模2除，得到余数R（n位，不足n位前面用0补齐）即为CRC校验码



CRC校验能力：能检测出少于n+1位的突发错误

接收端如何计算？

CRC计算示例

D = 1010001101

n = 5, G = 110101

R = 01110

实际传输数据：101000110101110

```
                      1101010110
G → 110101 | 101000110100000 ← 2^nD
             110101
              111011
              110101
                111010
                110101
                  111110
                  110101
                    101100
                    110101
                     110010
                     110101
                       01110 ← R
```

- 生成多项式G也可以表示成如下：

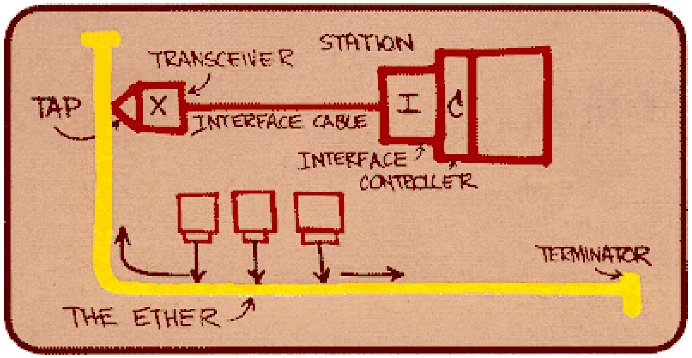
$$G = 1\cdot x^5 + 1\cdot x^4 + 0\cdot x^3 + 1\cdot x^2 + 0\cdot x^1 + 1\cdot x^0$$
$$= x^5 + x^4 + x^2 + x^0$$

- 例如，以太网的生成多项式为：

$$G = x^{32}+x^{26}+x^{23}+x^{22}+x^{16}+x^{12}+x^{11}+x^{10}+x^8+x^7+x^5+x^4+x^2+x+1$$

以太网的发展

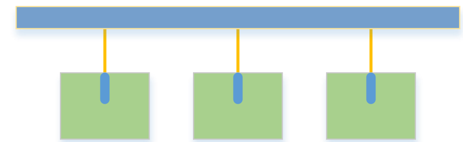
- 最初设计：1972年由施乐公司研制开发
- Internet标准：
  - 1980年：以太网（DIX v1.0）
  - **1982年：以太网II（DIX v2.0）**
- IEEE 802.3标准（1982年-今）
  - 10M以太网：1982年-1995年，主要支持同轴电缆、双绞线、光纤
  - 100M以太网：1995年-1998年，主要支持双绞线、光纤
  - 1000M以太网：1998年-，主要支持双绞线、光纤
  - 10G以太网：2002年-，主要支持双绞线、光纤
  - 40G以上以太网：2010年-，主要支持光纤

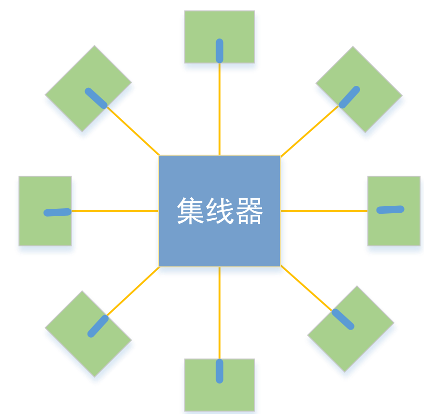


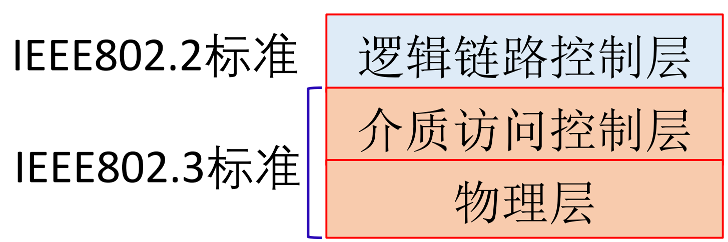
5.5 共享式以太网

共享式以太网连接方式与功能

- 连接方式
  - 方式1：通过同轴电缆连接（10Mbps以太网）
  - 方式2：通过双绞线、光纤和集线器连接
- 服务
  - 面向非连接的不可靠服务
- 功能
  - 物理层
    - 信号编码、时钟同步等，如差分曼彻斯特编码
  - 介质访问控制层
    - 介质访问控制
    - 差错检测
  - 逻辑链路控制层
    - 复用与分用







5.5 共享式以太网

IEEE 802.3帧结构

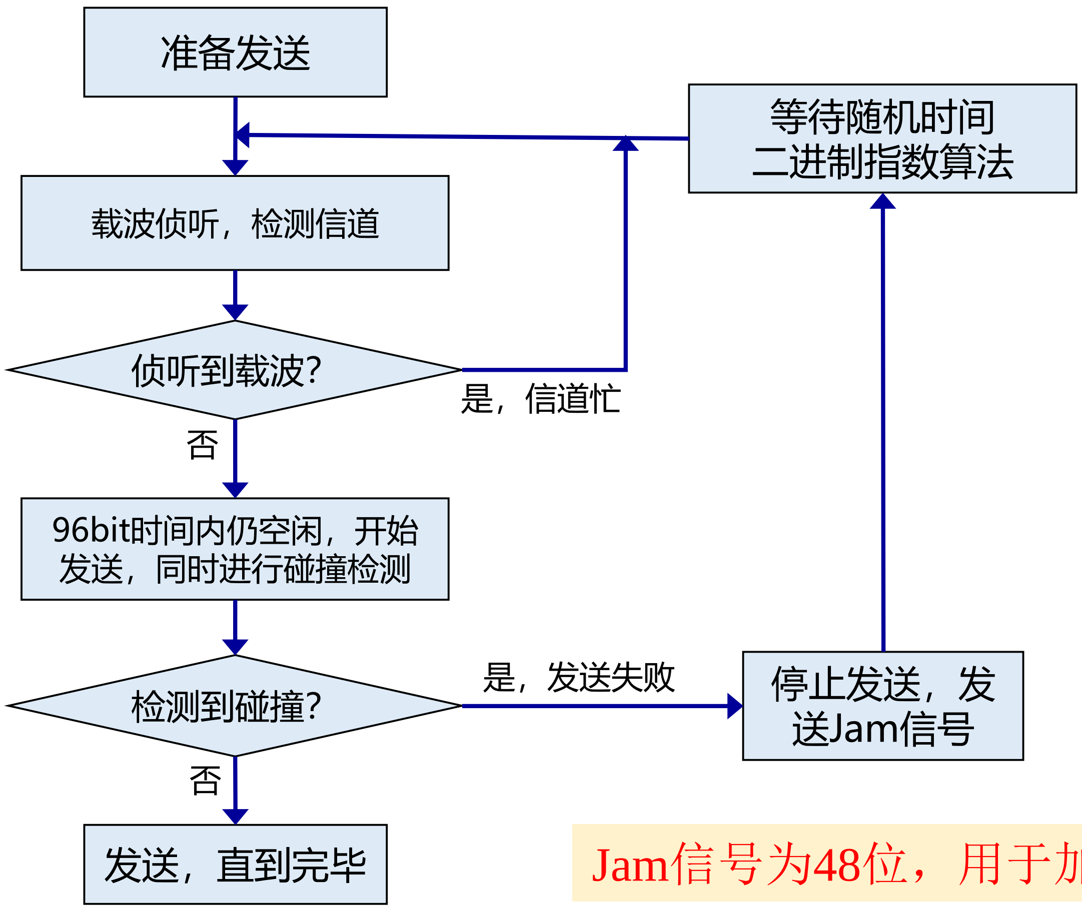
7octets	1	6	6	2	46-1500		4
Preamble	SFD	DA	SA	Length	LLC Data	Pad	CRC

DIX Ethernet V2帧结构

7octets	1	6	6	2	46-1500		4
Preamble	SFD	DA	SA	Type	Data	Pad	CRC

- 前导码: 位模式为10101010，用于时钟同步
- 帧开始定界符（SFD）: 位模式为10101011, 指明帧的开始
- CRC校验: $G=x^{32}+x^{26}+x^{23}+x^{22}+x^{16}+x^{12}+x^{11}+x^{10}+x^{8}+x^{7}+x^{5}+x^{4}+x^{2}+x+1$
- 类型：指明上层协议类型，
  - >= X0800
  - X0800　IP
  - X0806　ARP

■ 介质访问控制方法CSMA/CD



Jam信号为48位，用于加强冲突信号

CSMA/CD：Carrier Sense Multiple Access/ collisions detected

介质访问控制方法CSMA/CD

```
A: sense channel, if idle
    then {
        transmit and monitor the channel;
        If detect another transmission
          then {
              abort and send jam signal;
              update # collisions;
              delay as required by binary exponential backoff algorithm;
              goto A
              }
          else {done with the frame; set collisions to zero}
        }
    else {wait until ongoing transmission is over and goto A}
```

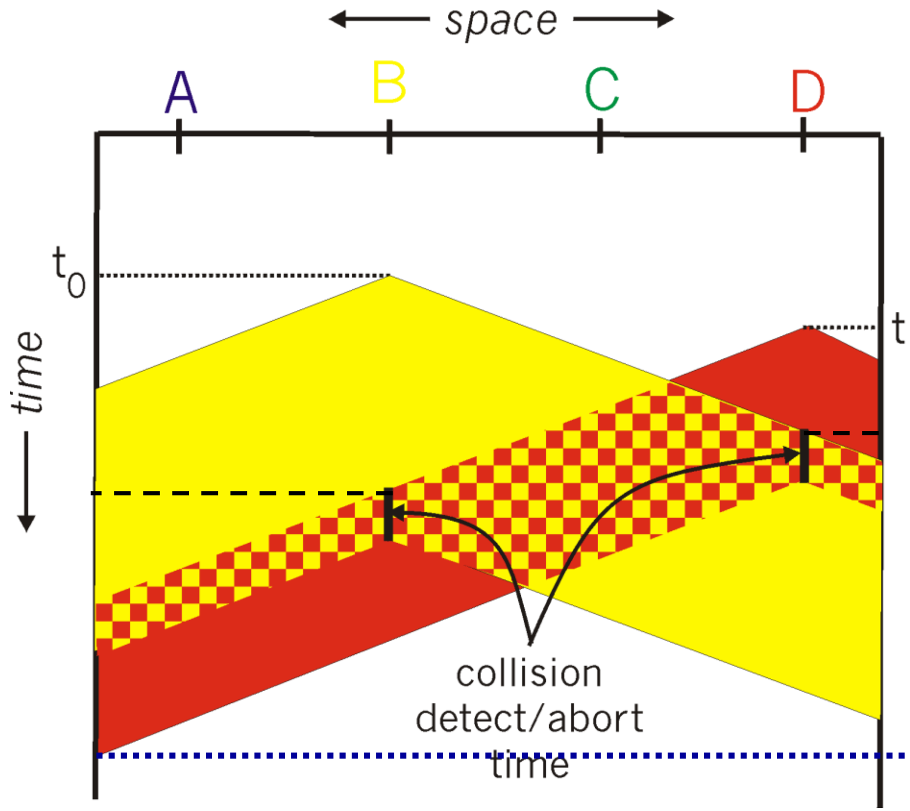
■ 二进制指数退后算法

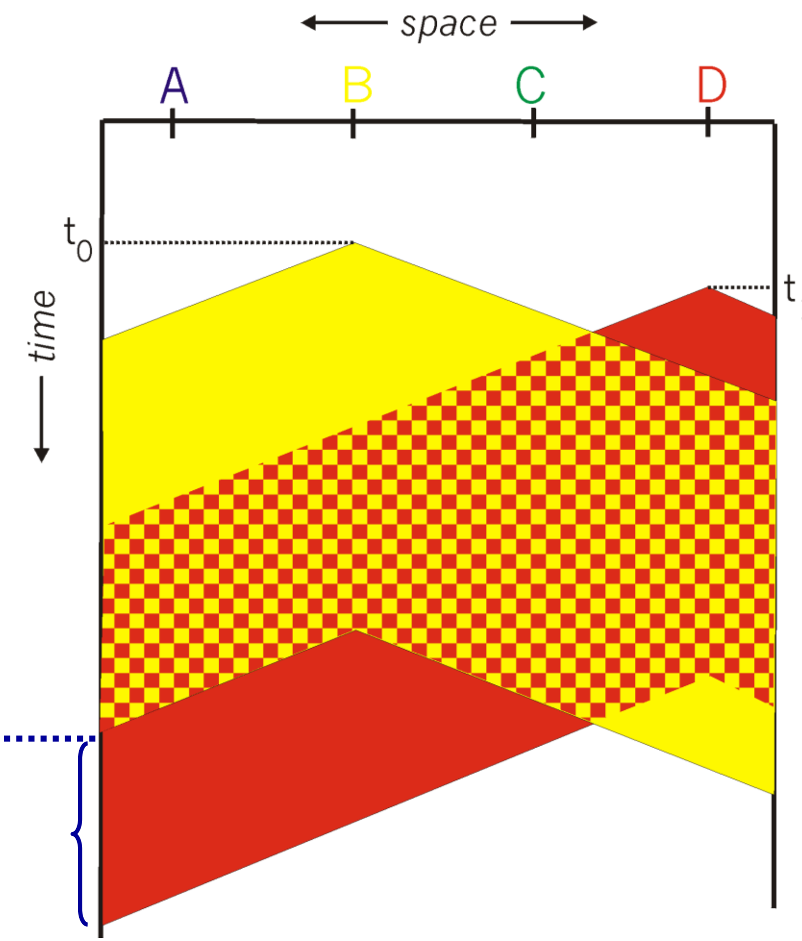
✓ 发生碰撞的站在停止发送数据后，要推迟（退避）一个随机时间才能再次发送数据。

✓ **目标**：根据当前负载情况，自适应地选择等待重试的时间：重负载，随机等待时间长；轻负载，随机等待时间短

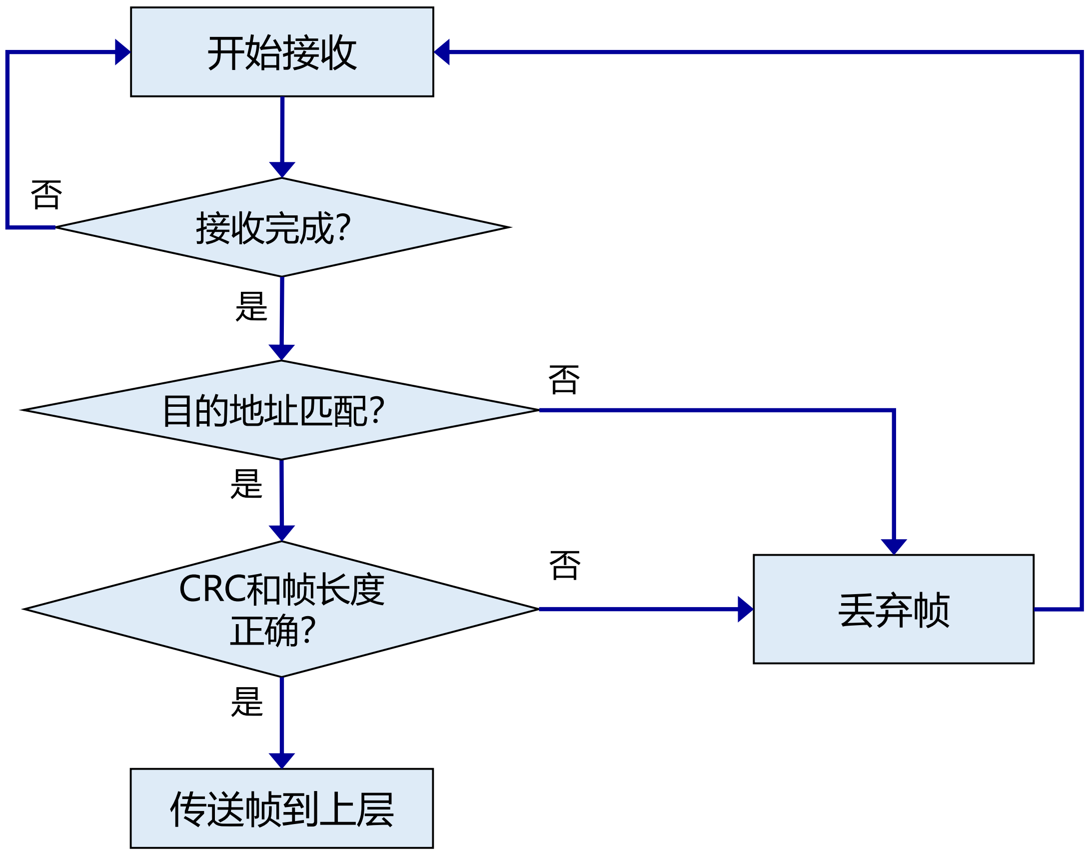
1. 基本退避时间取512位的传输时间，记为τ
2. 从整数集合 $\{0, 1, \ldots, (2^n-1)\}$中随机地取出一个数，记为$k$。重传所需的退后时间为 k 倍的基本退避时间，即$k\cdot\tau$ 。
3. 参数 n 的计算公式：$n = \min[$重传次数$, 10]$
4. 当 $n \le 10$ 时，参数 n 等于重传次数。
5. 当重传达 16 次仍不能成功时即丢弃该帧，并向高层报告。

CSMA/CD冲突检测


space
A
B
C
D
t0
t1
time
collision
detect/abort
time

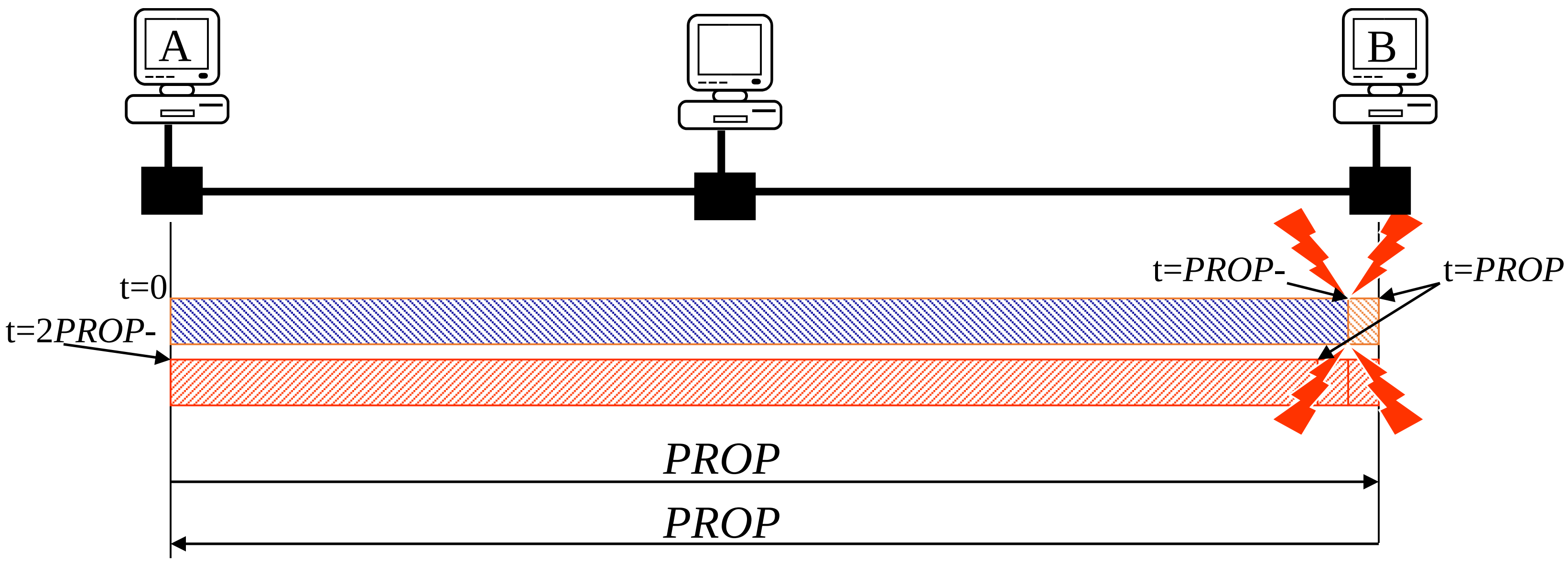

space
A
B
C
D
t0
t1
time

以太帧的接收



■ CSMA/CD对网络规模的约束

为了确保帧在传输过程中没有冲突，节点需要在完成传输之前检测到冲突



Events:

t=0: Host A starts transmitting a packet.

t=*PROP*--: Just before the first bit reaches Host B, Host B senses the line to be idle and starts to transmit a packet.

t=*PROP*-: A collision takes place near Host B.

t=*PROP*: Host B receives data whilst transmitting, and so detects the collision.

t=2*PROP*-: Host A receives data whilst transmitting, and so detects the collision.

5.5 共享式以太网

■ 速率增长带来的问题

- ✓ 速率增长：10Mb/s →100Mb/s →1Gb/s →10Gb/s
- ✓ 问题: $TRANSP > 2PROP$
- ✓ 例如：如果在2500米的电缆上以100Mb/s速率运行 CSMA/CD

$$PROP_{max} = l/c = 2500/(2.5*10^8)=10\ \mu s$$

$$TRANSP > 2PROP \ \ => TRANSP > 20\ \mu s$$

$$Packetsize \geq 20\mu s * 100Mb/s = 2000 bits$$

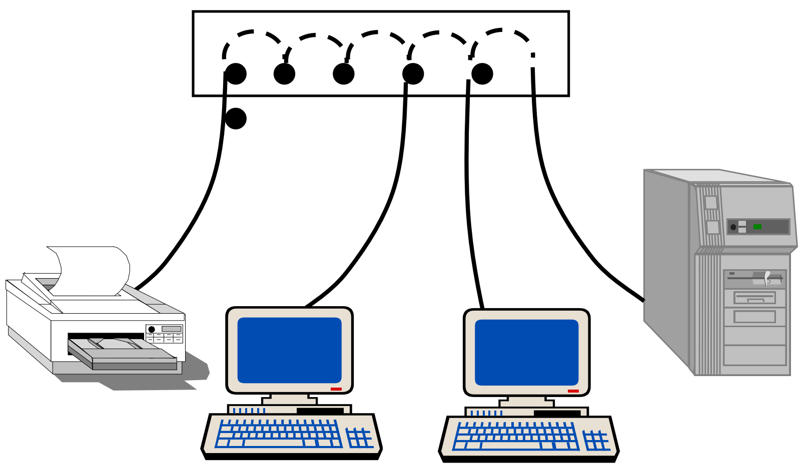
如果要保持最小帧长度不变，最大传输距离如何限制？

5.6 交换式以太网

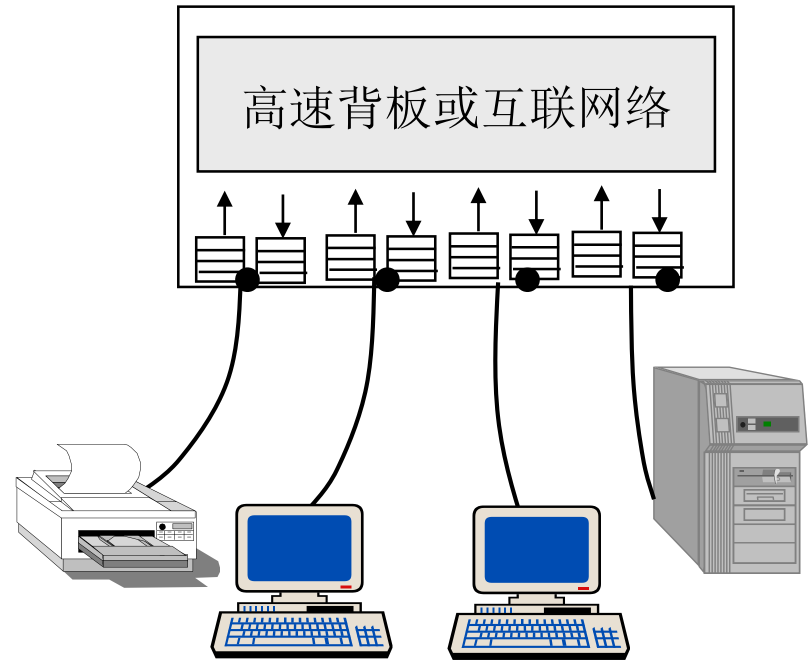
集线器与交换机

集线器

广播式传输



交换机

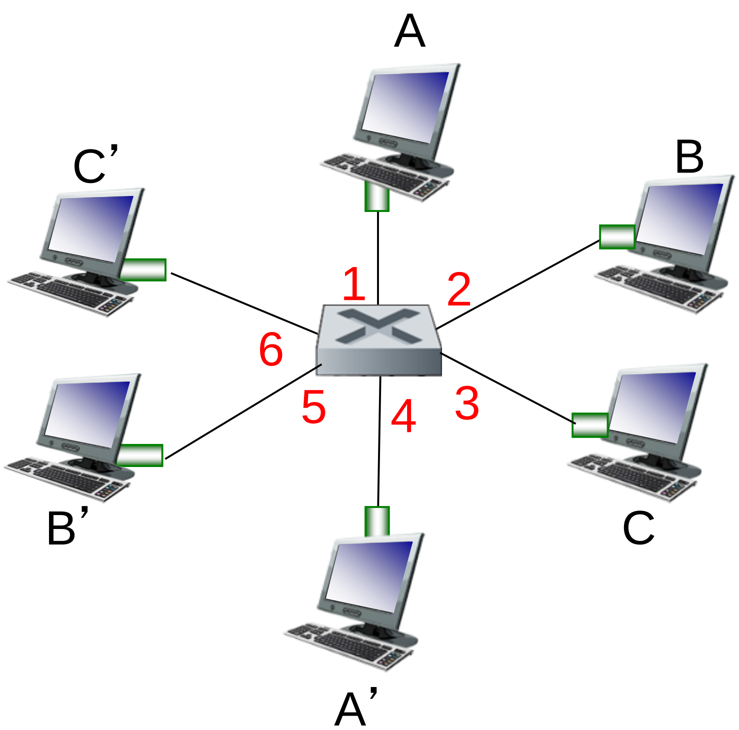


Ethernet switch

- link-layer device
  - store, forward Ethernet frames
  - examine incoming frame's MAC address, selectively forward frame to one-or-more outgoing links when frame is to be forwarded on segment, uses CSMA/CD to access segment
- *transparent*
  - nodes are unaware of presence of switches
- *plug-and-play, self-learning*
  - switches do not need to be configured

Switch: *multiple* simultaneous transmissions

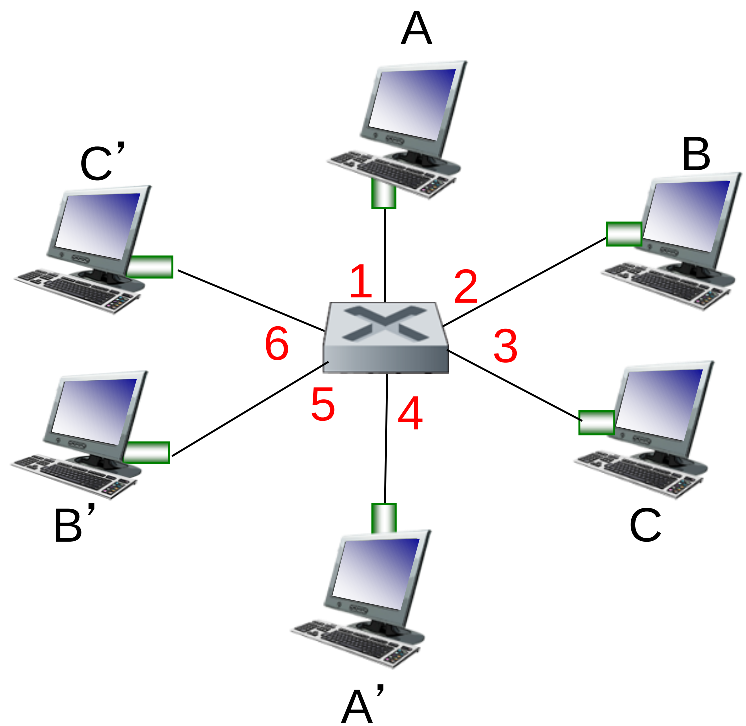
- nodes have dedicated, direct connection to switch
- switches buffer packets
- Ethernet protocol used on *each* link, but no collisions; full duplex
- *switching:* A-to-A' and B-to-B' can transmit simultaneously, without collisions



switch with six interfaces (1,2,3,4,5,6)

Switch Table

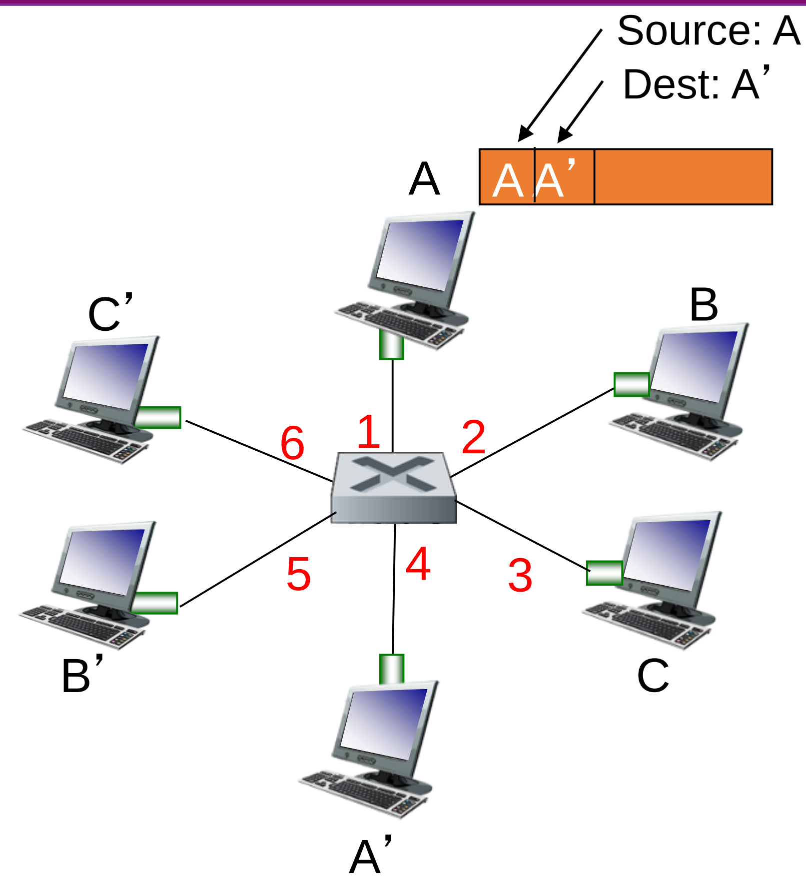
- *Q:* how does switch know that A' reachable via interface 4, B' reachable via interface 5?
- *A:* each switch has a switch table, each entry:
  - (MAC address of host, interface to reach host, time stamp)
- looks like a routing table!
- *Q:* how are entries created, maintained in switch table?
  - something like a routing protocol?



switch with six interfaces (1,2,3,4,5,6)

Switch: self-learning

- switch *learns* which hosts can be reached through which interfaces
  - when frame received, switch "learns" location of sender: incoming LAN segment
  - records sender/location pair in switch table



MAC addr	interface	TTL
A	1	60

Switch table (initially empty)

Switch: frame filtering/forwarding

When frame received:

1. record link associated with sending host
2. index switch table using MAC dest. address
3. **if** entry found for destination
   then {

   if dest. on segment from which frame arrived
   then drop the frame

   else forward the frame on interface indicated

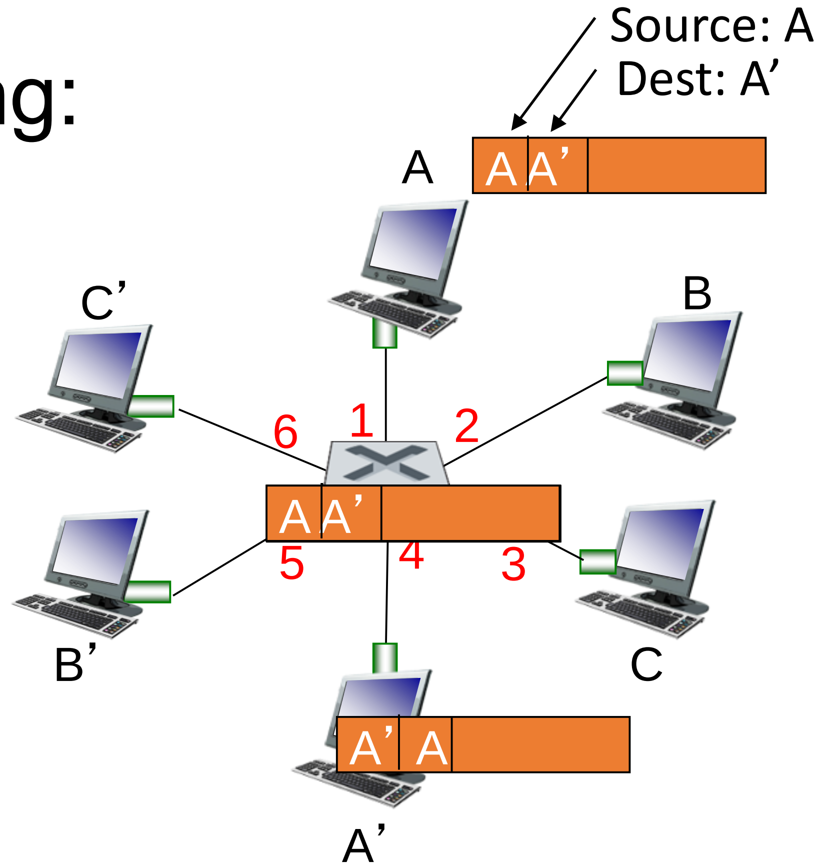
   }

   else flood

forward on all but the interface on which the frame arrived

Self-learning, forwarding: example

- frame destination unknown: flood
- destination A location known: selective send

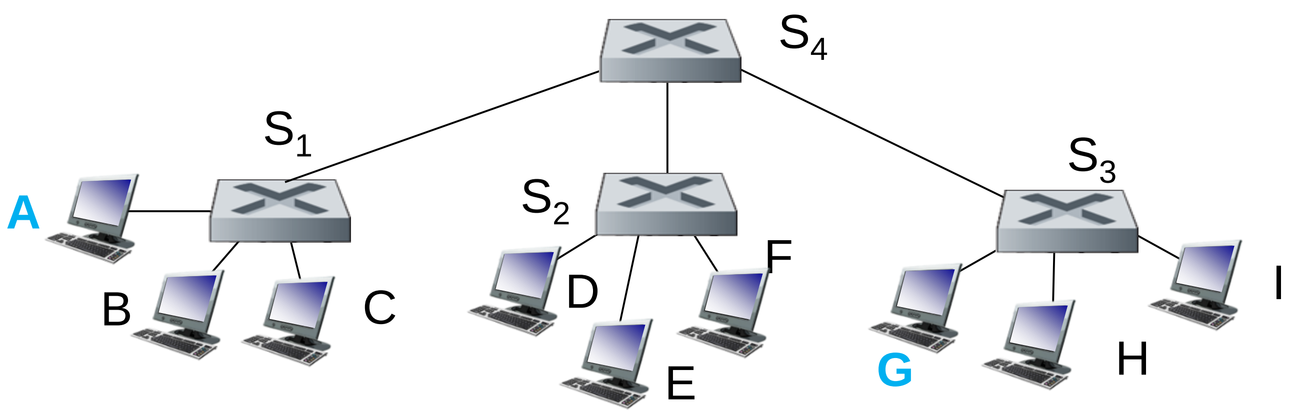


MAC addr	interface	TTL
A	1	60
A'	4	60

Switch table (initially empty)

Interconnecting switches

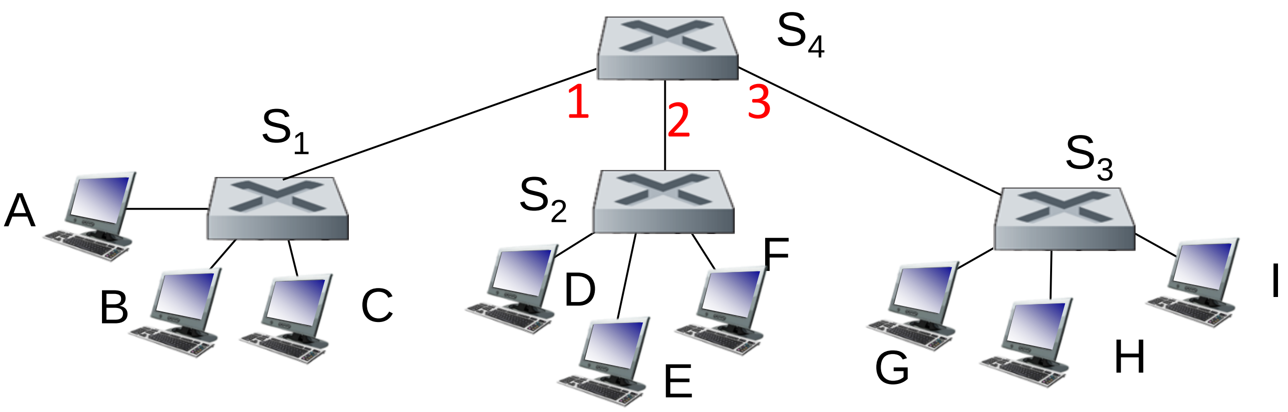
- switches can be connected together



- Q: sending from A to G - how does S_1 know to forward frame destined to G via S_4 and S_3?
- A: self learning! (works exactly the same as in single-switch case!)

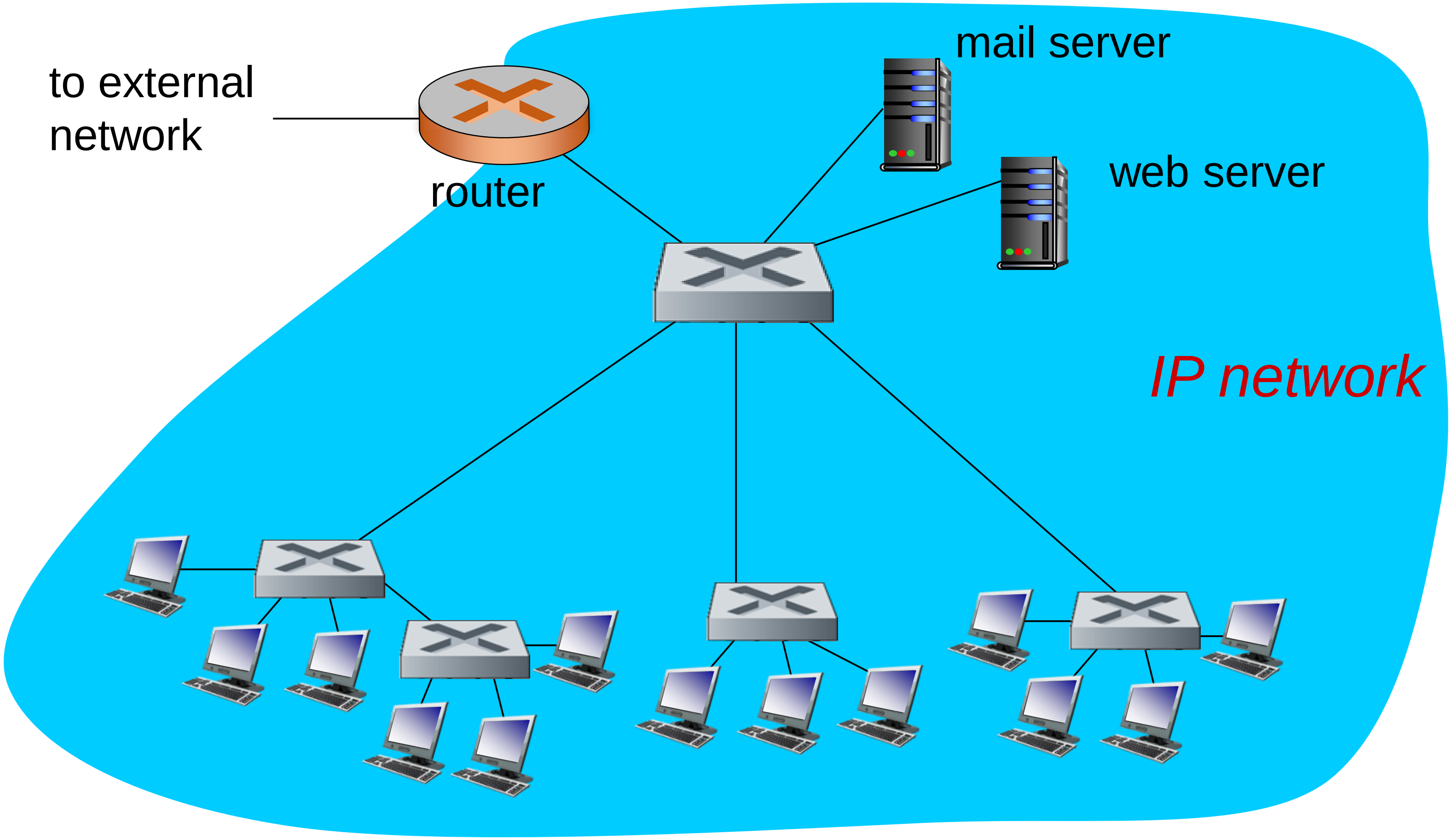
Self-learning multi-switch example

- Suppose C sends frame to I, I responds to C



- Q: show switch tables and packet forwarding in S_1, S_2, S_3, S_4

Institutional network



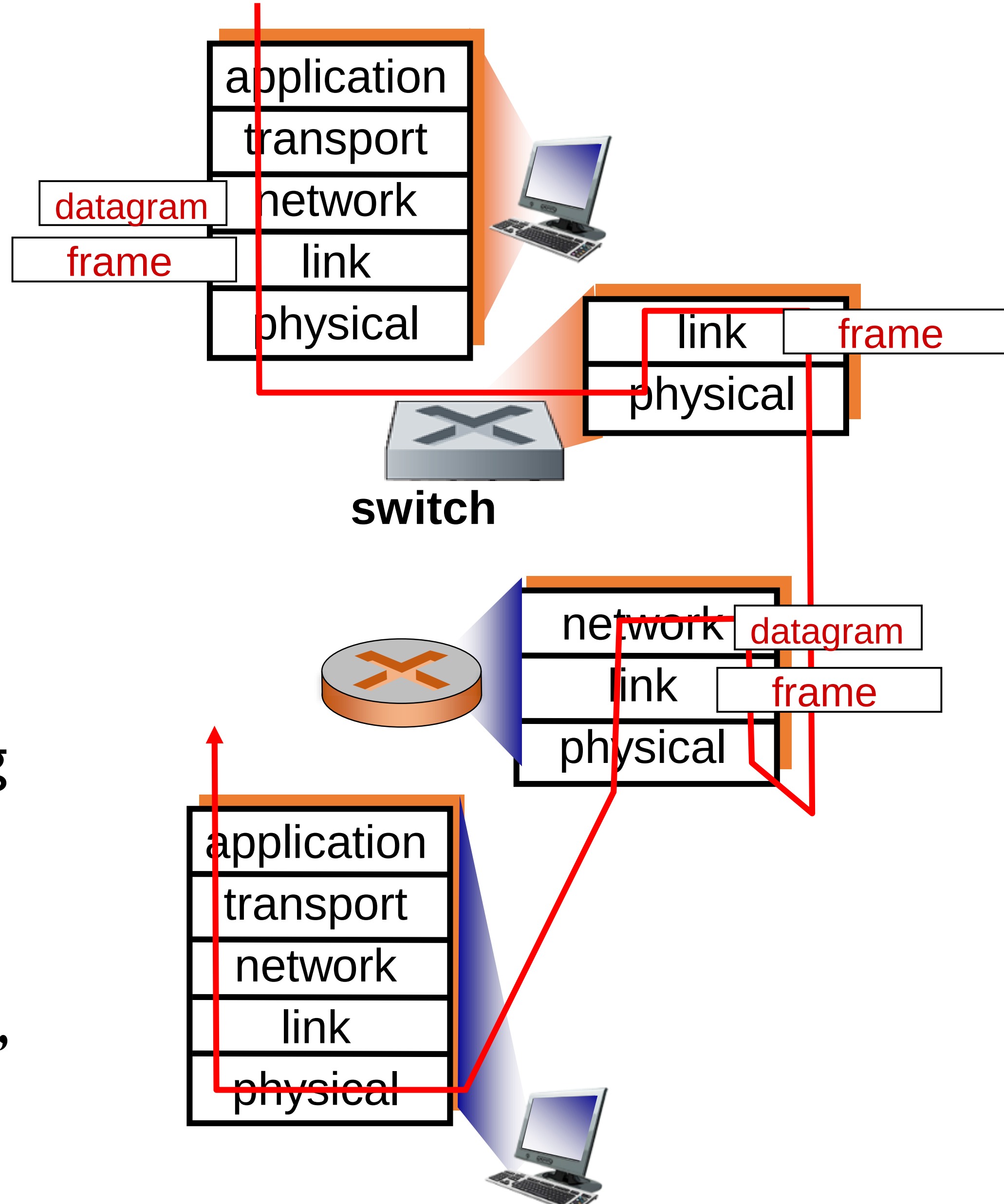
Switches vs. routers

both are store-and-forward:

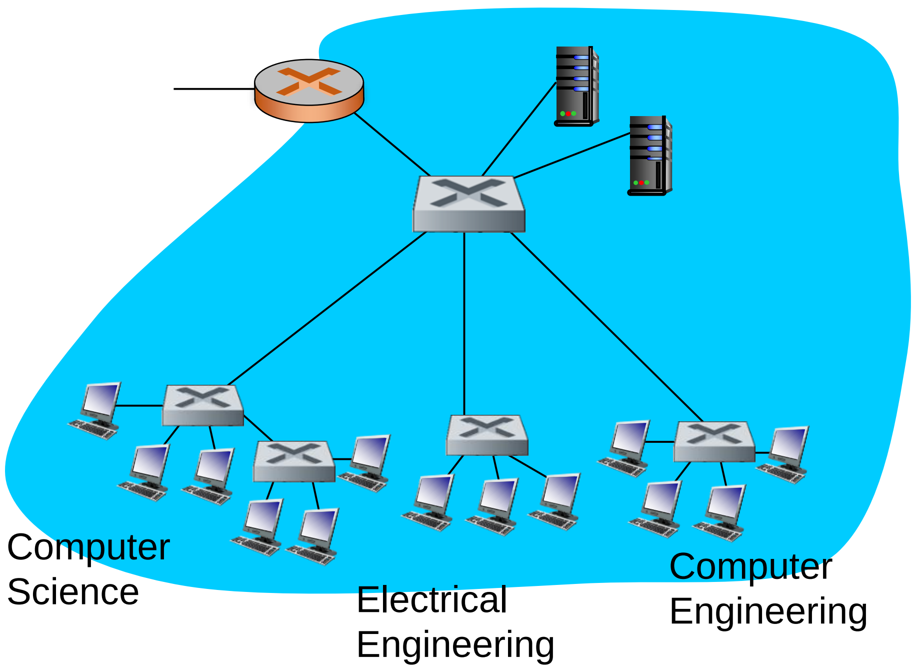
- *routers:* network-layer devices (examine network-layer headers)
- *switches*: link-layer devices (examine link-layer headers)

both have forwarding tables:

- *routers:* compute tables using routing algorithms, IP addresses
- *switches:* learn forwarding table using flooding, learning, MAC addresses



VLANs: motivation



consider:

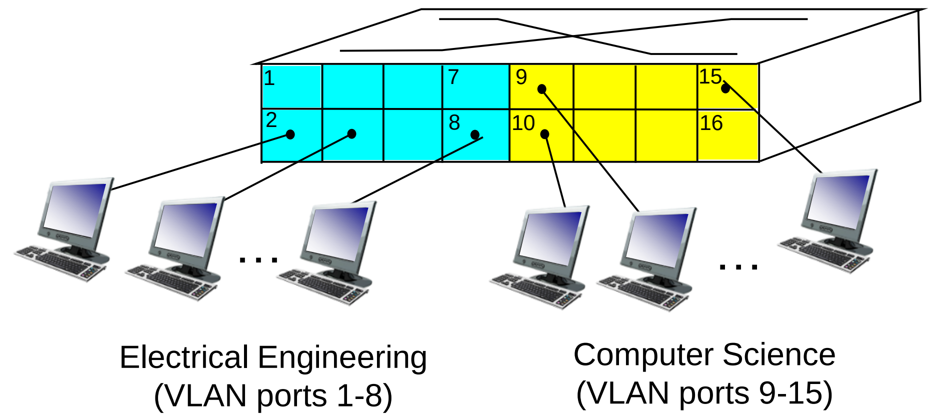
- CS user moves office to EE, but wants connect to CS switch?
- single broadcast domain:
  - all layer-2 broadcast traffic (ARP, DHCP, unknown location of destination MAC address) must cross entire LAN
  - security/privacy, efficiency issues

VLANs

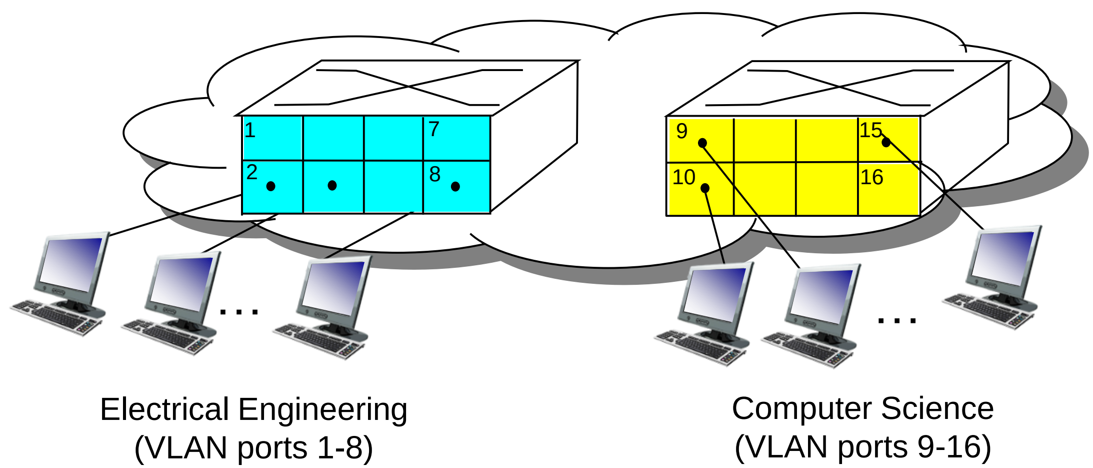
Virtual Local Area Network

switch(es) supporting VLAN capabilities can be configured to define multiple ***virtual*** LANS over single physical LAN infrastructure.

port-based VLAN: switch ports grouped (by switch management software) so that *single* physical switch ……

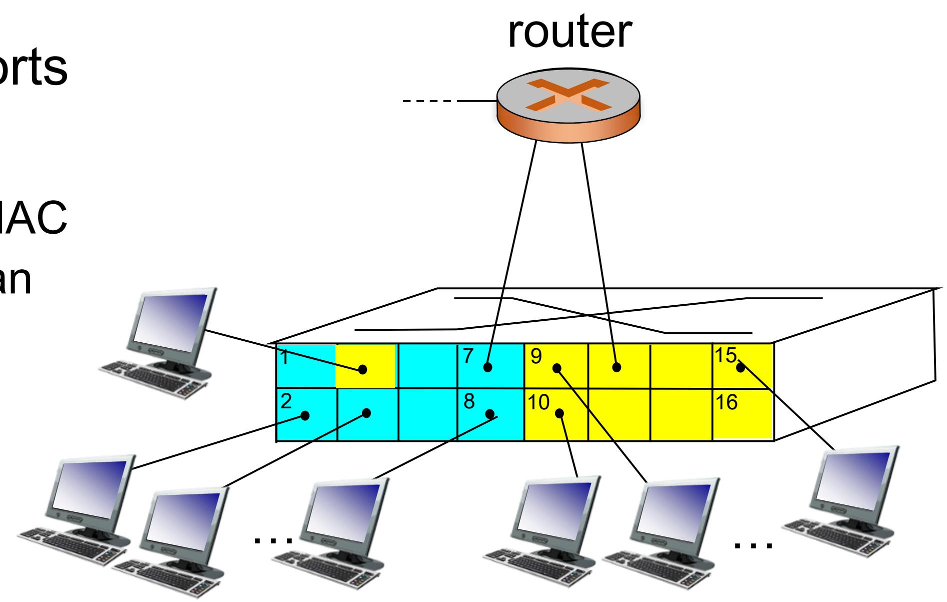


… operates as multiple virtual switches



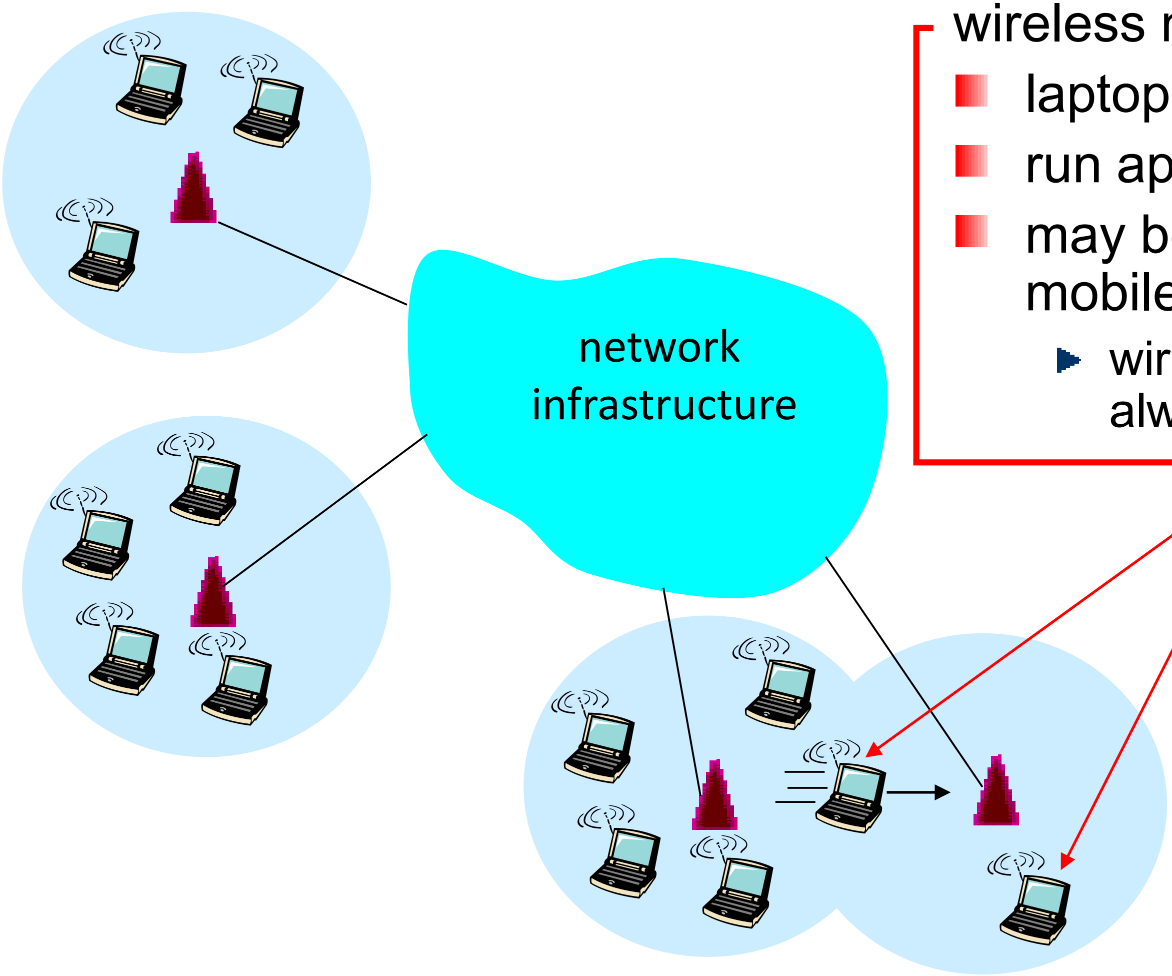
Port-based VLAN

- *traffic isolation:* frames to/from ports 1-8 can *only* reach ports 1-8
  - can also define VLAN based on MAC addresses of endpoints, rather than switch port
- dynamic membership: ports can be dynamically assigned among VLANs
- forwarding between VLANS: done via routing (just as with separate switches)
  - ✓ in practice vendors sell combined switches plus routers



三层交换技术？

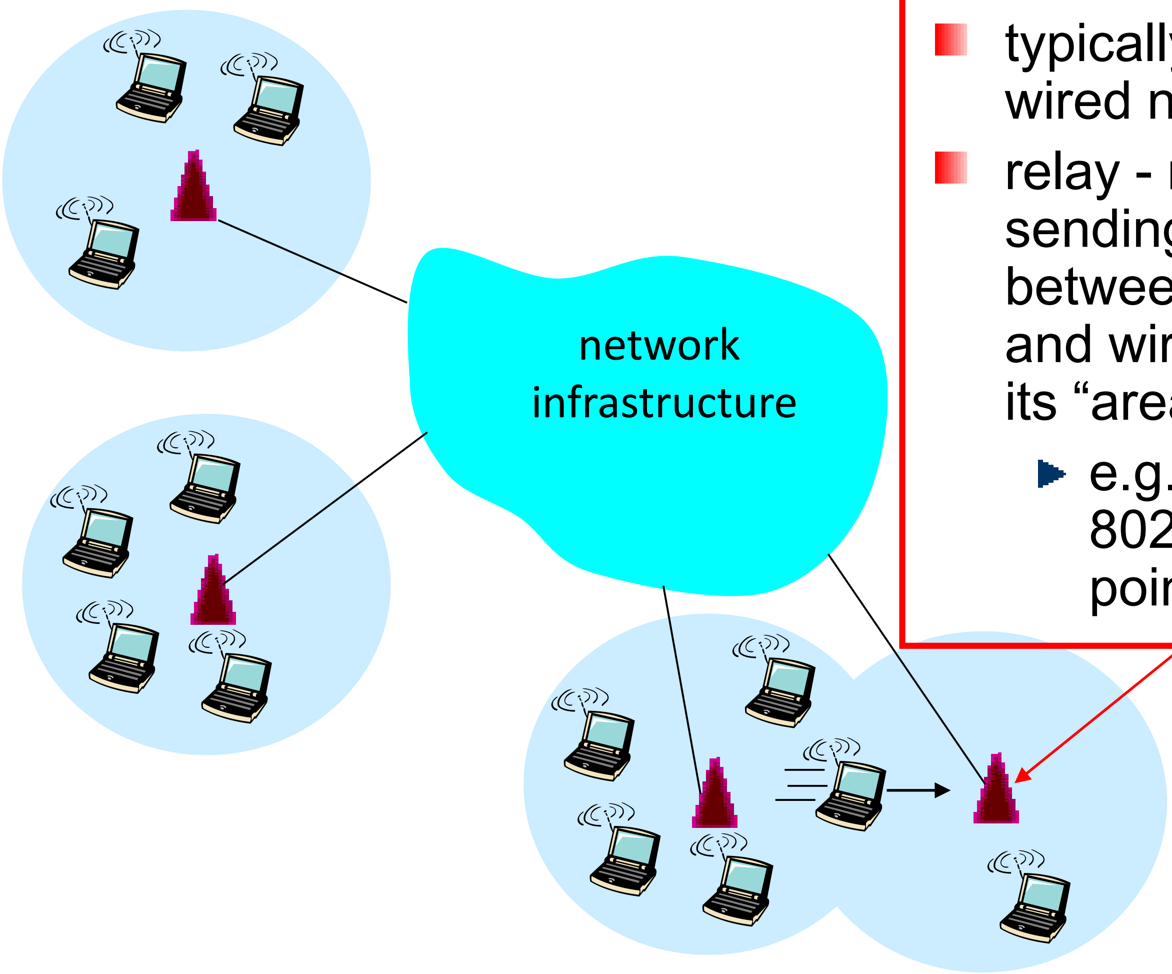
无线局域网构成元素



wireless nodes

- laptop, PDA, IP phone
- run applications
- may be stationary (non-mobile) or mobile
  - wireless does *not* always mean mobility

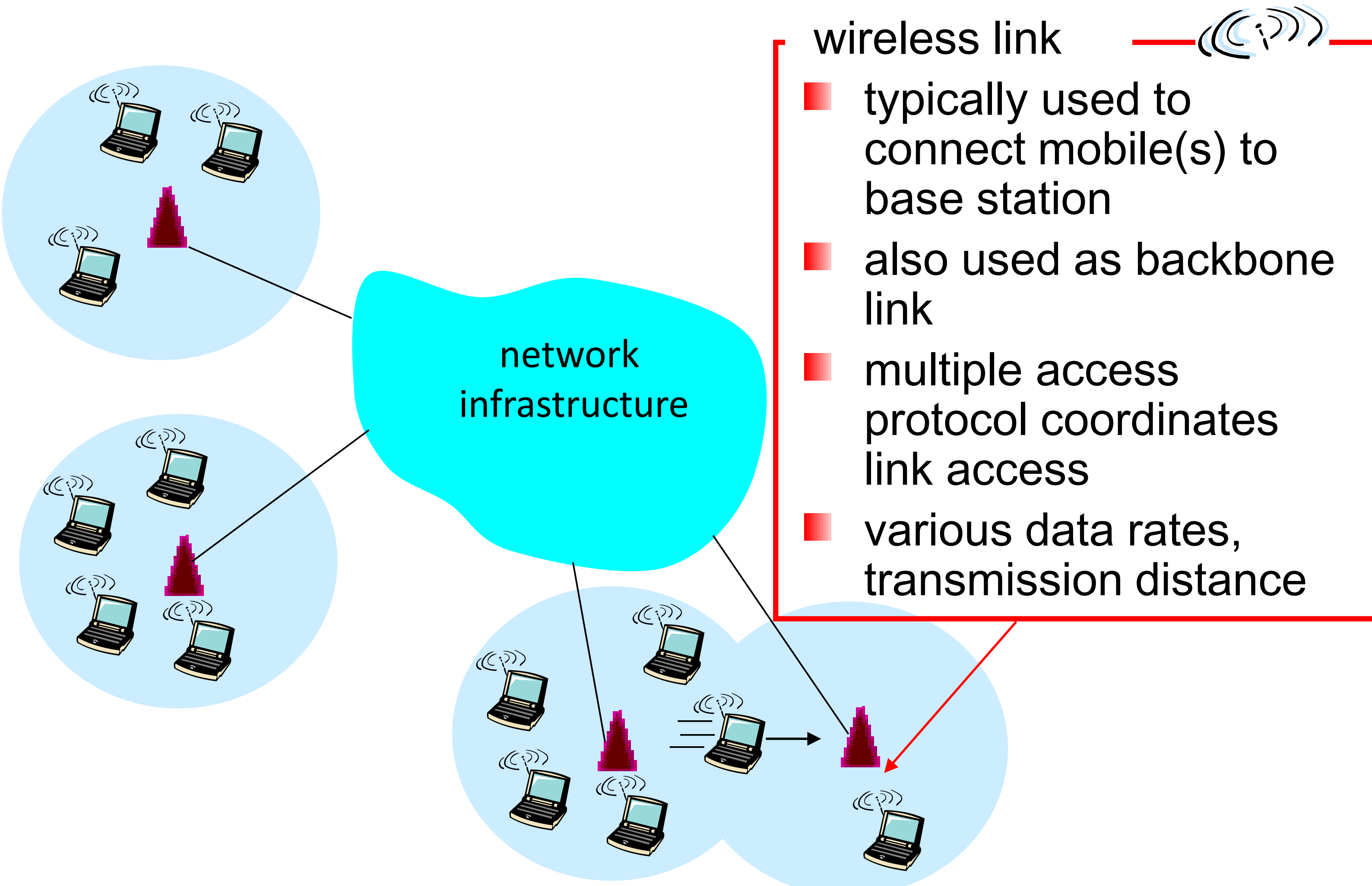
无线局域网构成元素



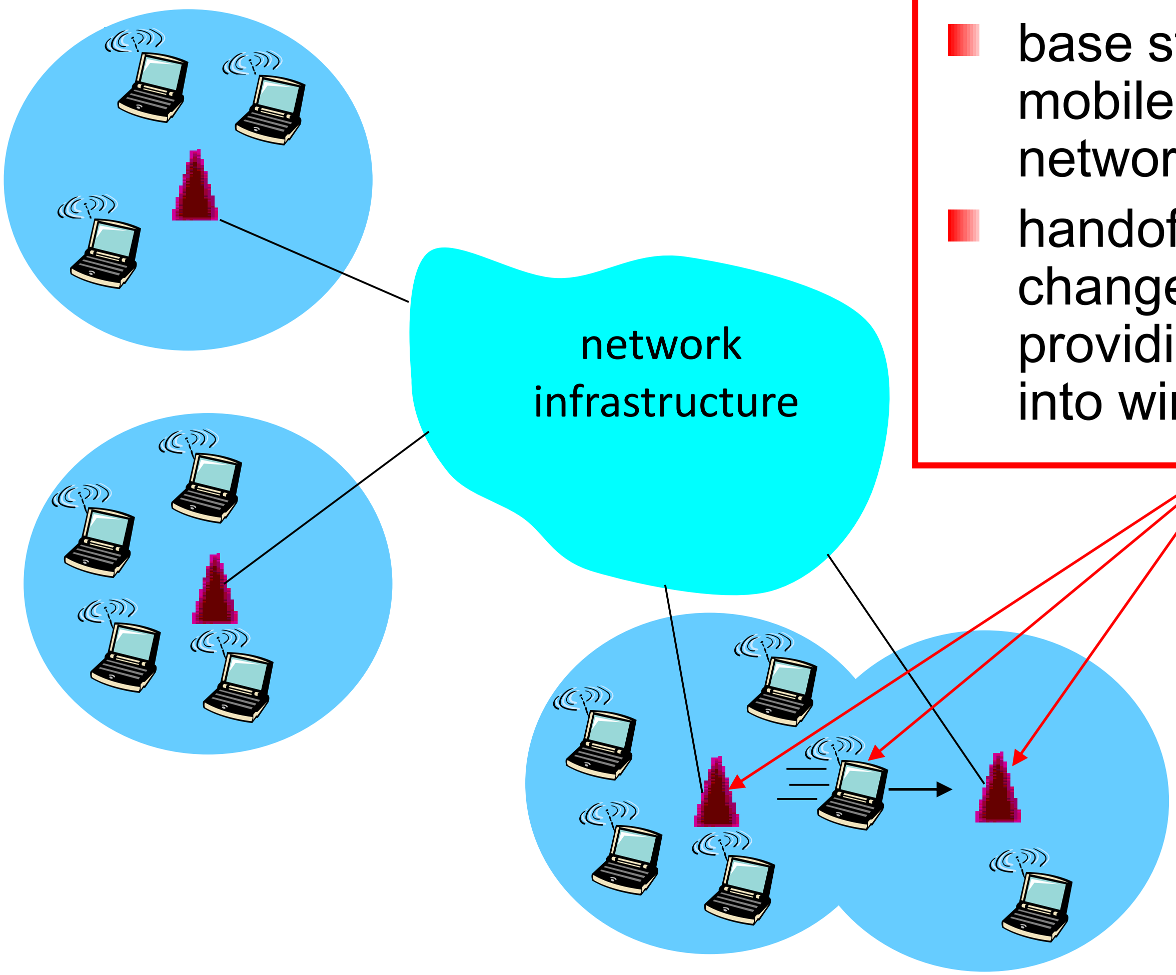
base station

- typically connected to wired network
- relay - responsible for sending packets between wired network and wireless node(s) in its “area”
  - e.g., cell towers 802.11 access points

无线局域网构成元素


network
infrastructure
wireless link
typically used to connect mobile(s) to base station
also used as backbone link
multiple access protocol coordinates link access
various data rates, transmission distance

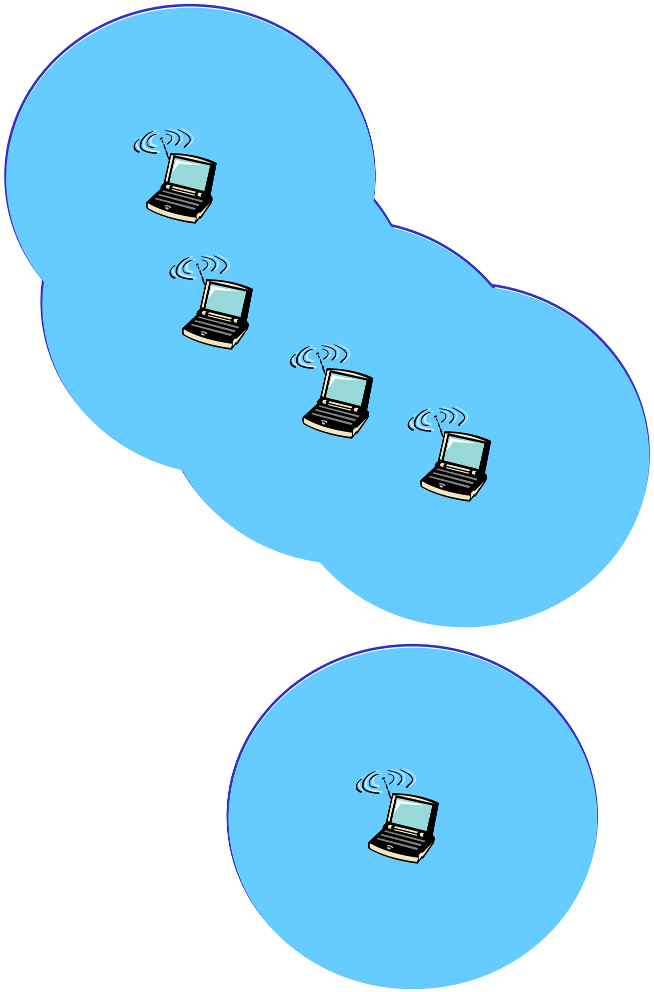
无线网络组网模式



infrastructure mode

- base station connects mobiles into wired network
- handoff: mobile changes base station providing connection into wired network

无线网络组网模式



Ad hoc mode

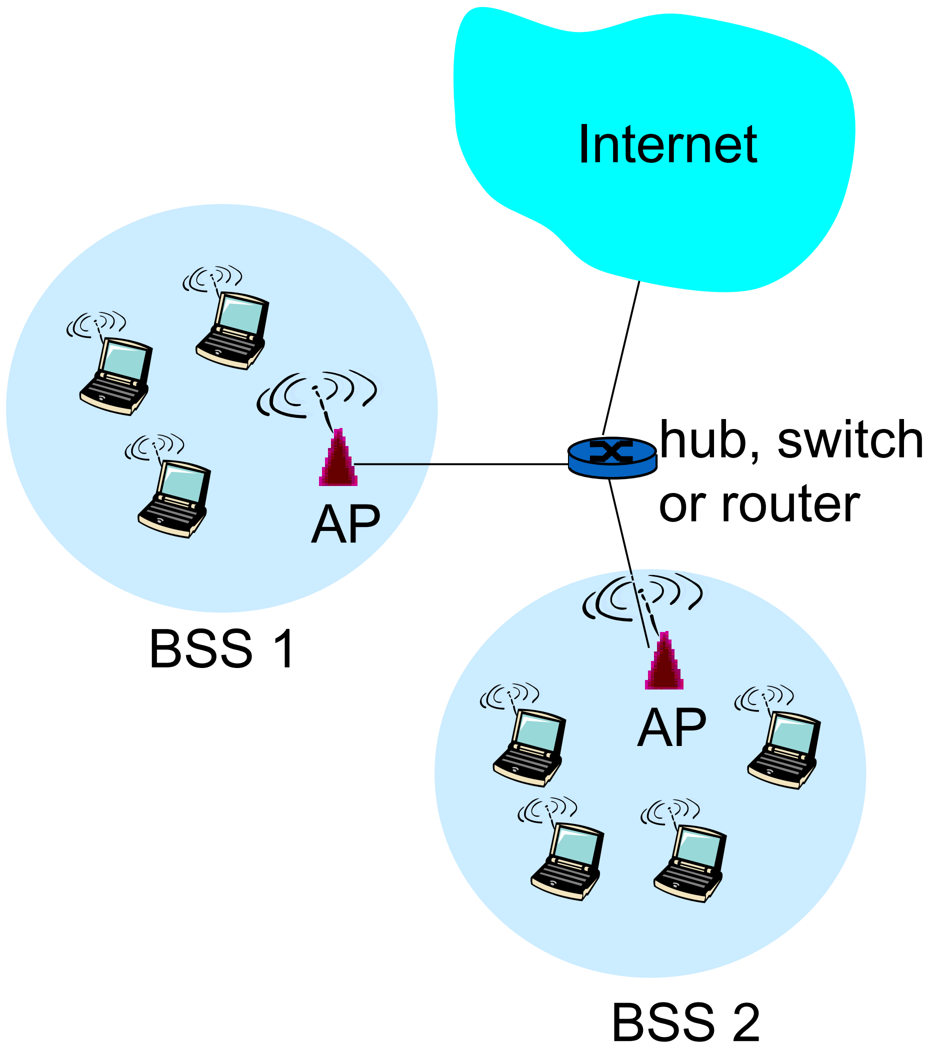
- no base stations
- nodes can only transmit to other nodes within link coverage
- nodes organize themselves into a network: route among themselves

IEEE 802.11无线局域网（WiFi）

- 802.11b
  - 2.4-2.485 GHz
  - up to 11 Mbps
- 802.11a
  - 5.1-5.8 GHz
  - up to 54 Mbps
- 802.11g
  - 2.4-2.485 GHz
  - up to 54 Mbps
- 802.11n:
  - 2.4-5 GHz
  - up to 200 Mbps
- 802.11ac
  - 5.1-5.8 GHz
  - up to 1.73Gbps

- All use CSMA/CA for multiple access
- All have base-station and ad-hoc network versions

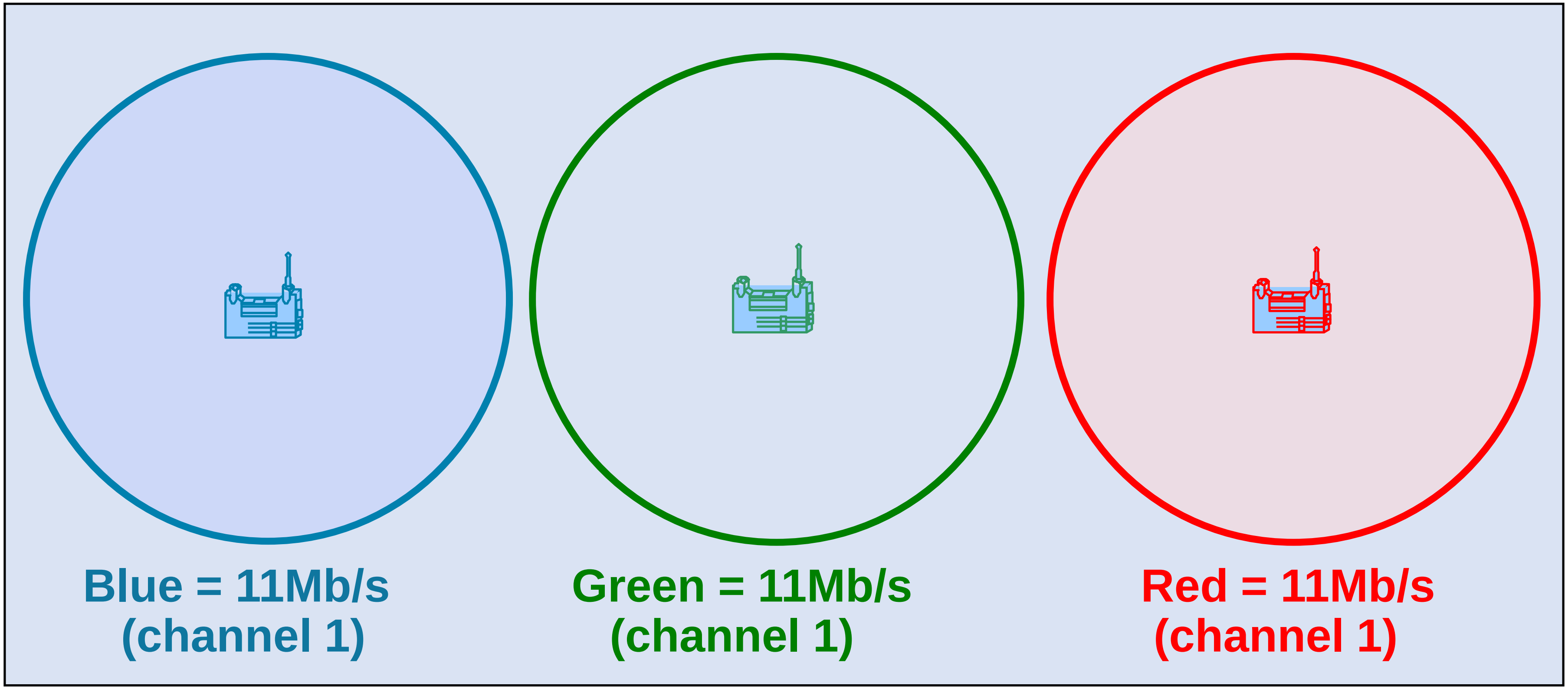
802.11 LAN结构



- Basic Service Set (BSS) in infrastructure mode contains:
  - wireless nodes
  - access point (AP): base station
  - wireless node communicates with base station
- Ad Hoc mode: nodes only
  - Independent Basic Service Set, IBSS

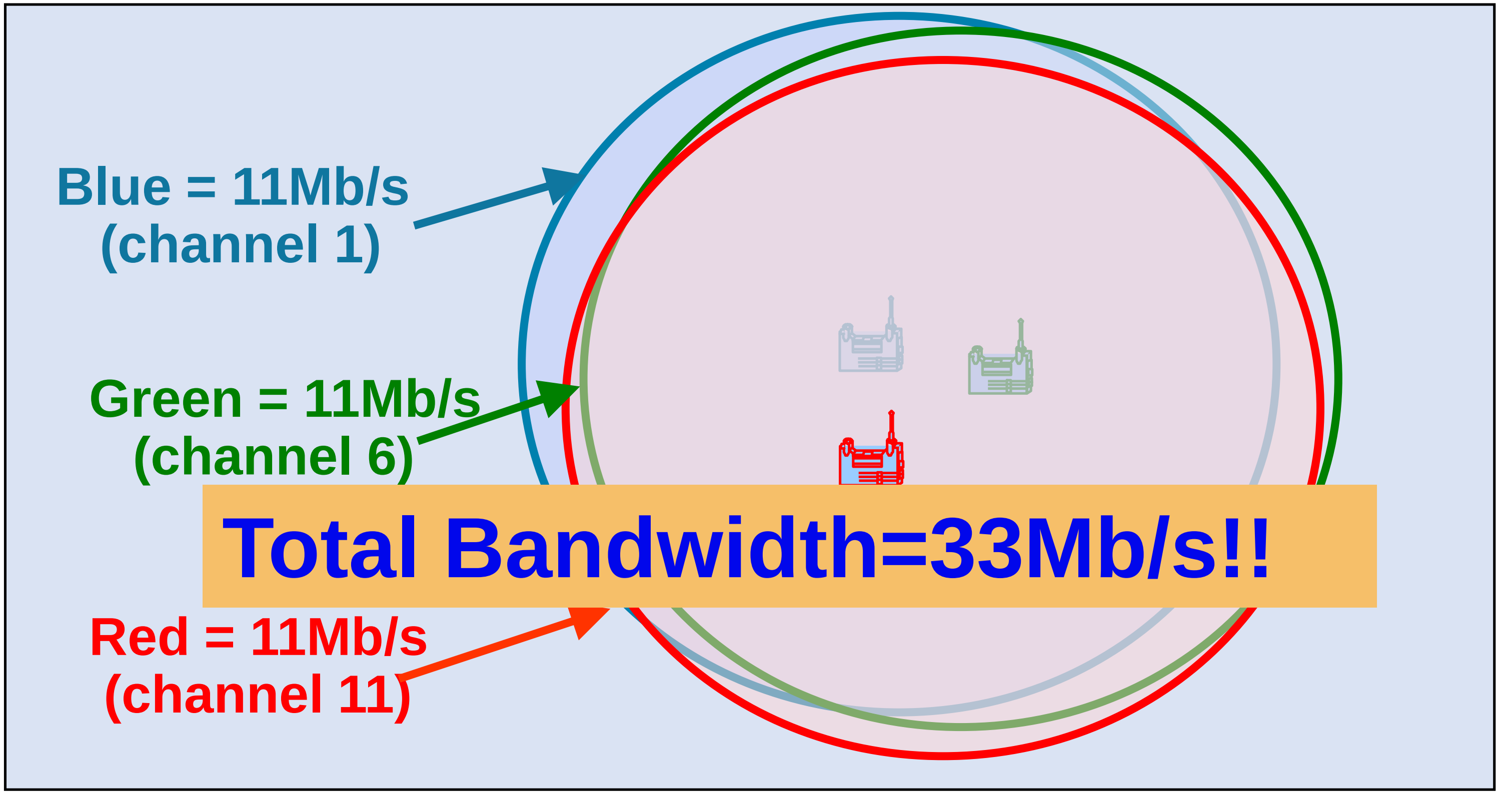
802.11b频谱空间重用

- 限制发射功率



802.11b频谱通道划分

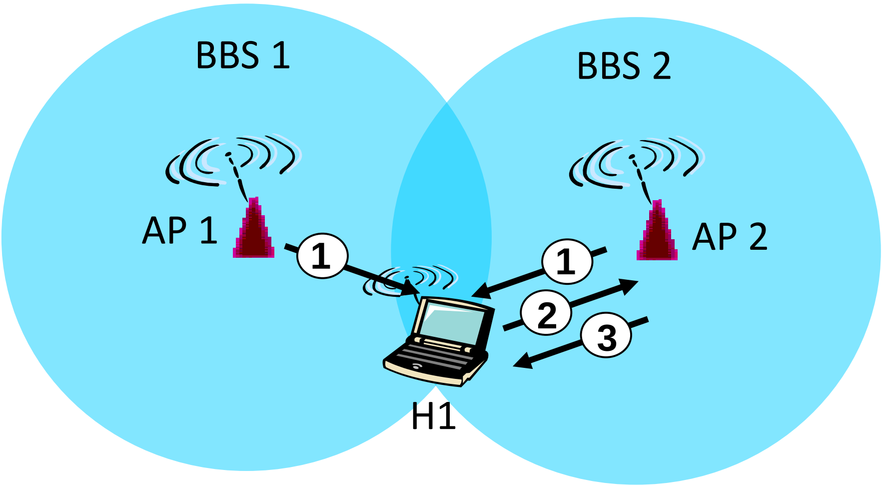
- 83.5MHz频带，划分11个通道（Channel）
- 扩频后每个通道22MHz，最多3个无干扰通道

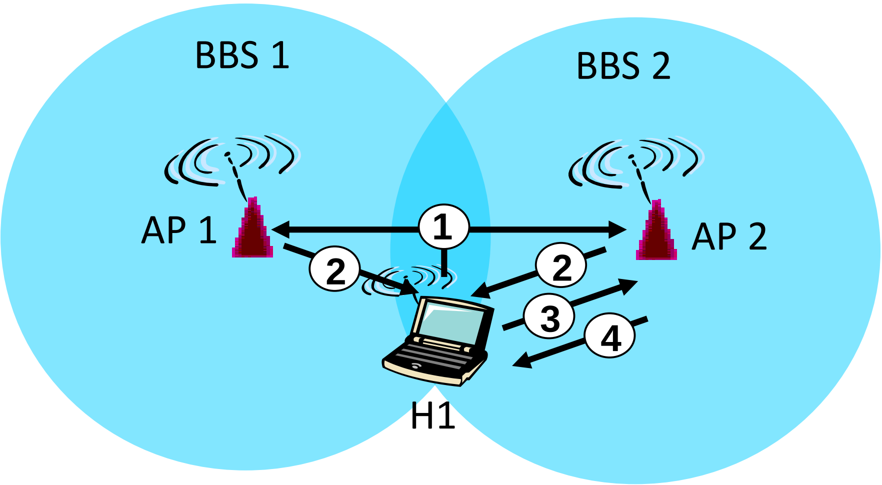


关联到AP

- Passive Scanning
  - scans channels, listening for *beacon frames* containing AP's name (SSID) and MAC address
  - selects AP to associate with
- Active Scanning
  - send a Probe Request frame
  - receive Probe Response frames
- may perform authentication (username+password/MAC address)
- will typically run DHCP to get IP address in AP's subnet

被动和主动扫描





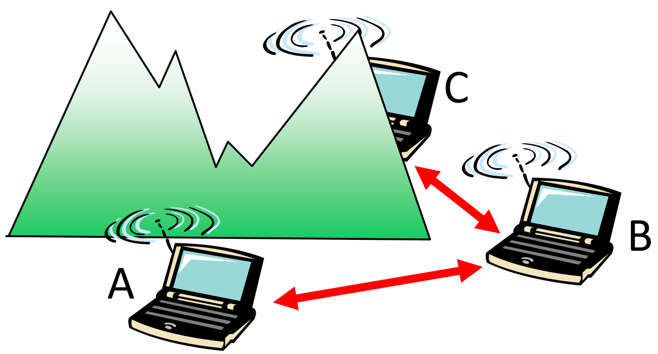
Passive Scanning:

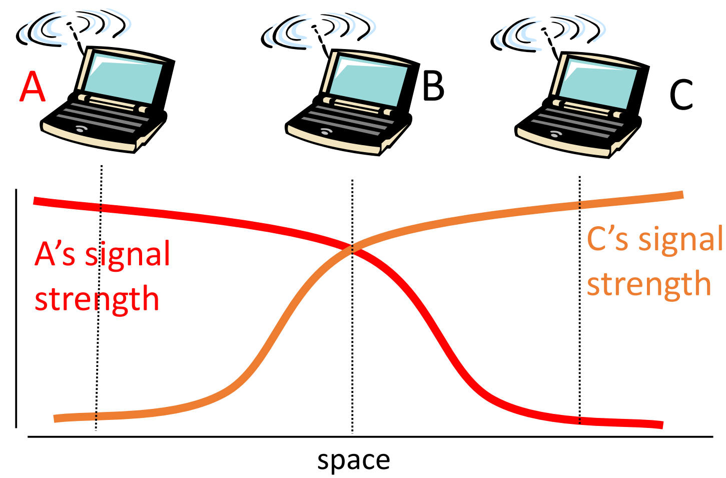
- ✓ Beacon frames sent from APs
- ✓ Association Request frame sent: H1 to selected AP
- ✓ Association Response frame sent: H1 to selected AP

Active Scanning:

- ✓ Probe Request frame broadcast from H1
- ✓ Probes response frame sent from APs
- ✓ Association Request frame sent: H1 to selected AP
- ✓ Association Response frame sent: H1 to selected AP

无线链路的特殊问题





Hidden terminal problem

- B, A hear each other
- B, C hear each other
- A, C can not hear each other
- means A, C unaware of their interference at B

Signal fading:

- B, A hear each other
- B, C hear each other
- A, C can not hear each other interfering at B

介质访问控制方法: CSMA/CA

- 802.11: CSMA - sense before transmitting
  - don’t collide with ongoing transmission by other node
- 802.11: no collision detection!
  - difficult to receive (sense collisions) when transmitting due to weak received signals (fading)
  - can’t sense all collisions in any case: hidden terminal, fading
  - goal: *avoid collisions:* CSMA/C(ollision)A(voidance)

介质访问控制方法: CSMA/CA

802.11 sender

1 if sense channel idle for **DIFS** then

transmit entire frame (no CD)

2 if sense channel busy then

start random backoff time

timer counts down while channel idle,

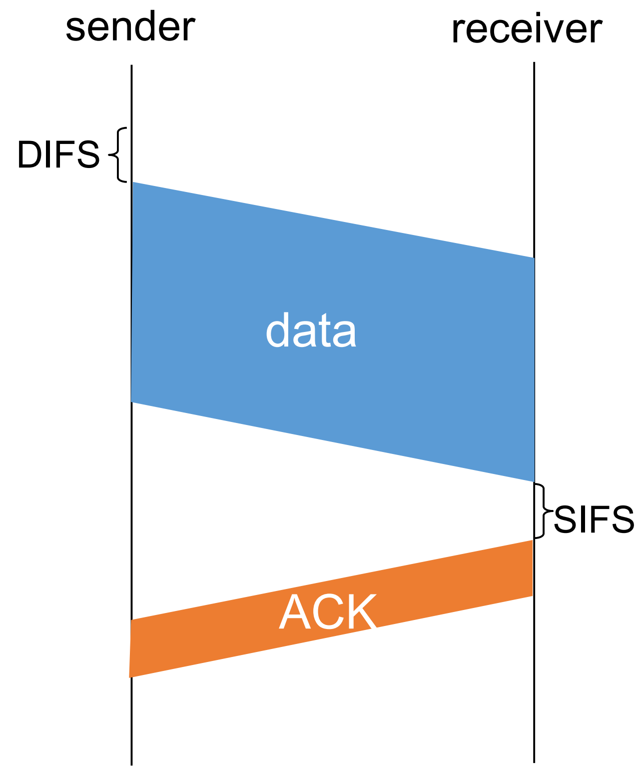
transmit when timer expires

3 if no ACK, increase random backoff interval, repeat 2

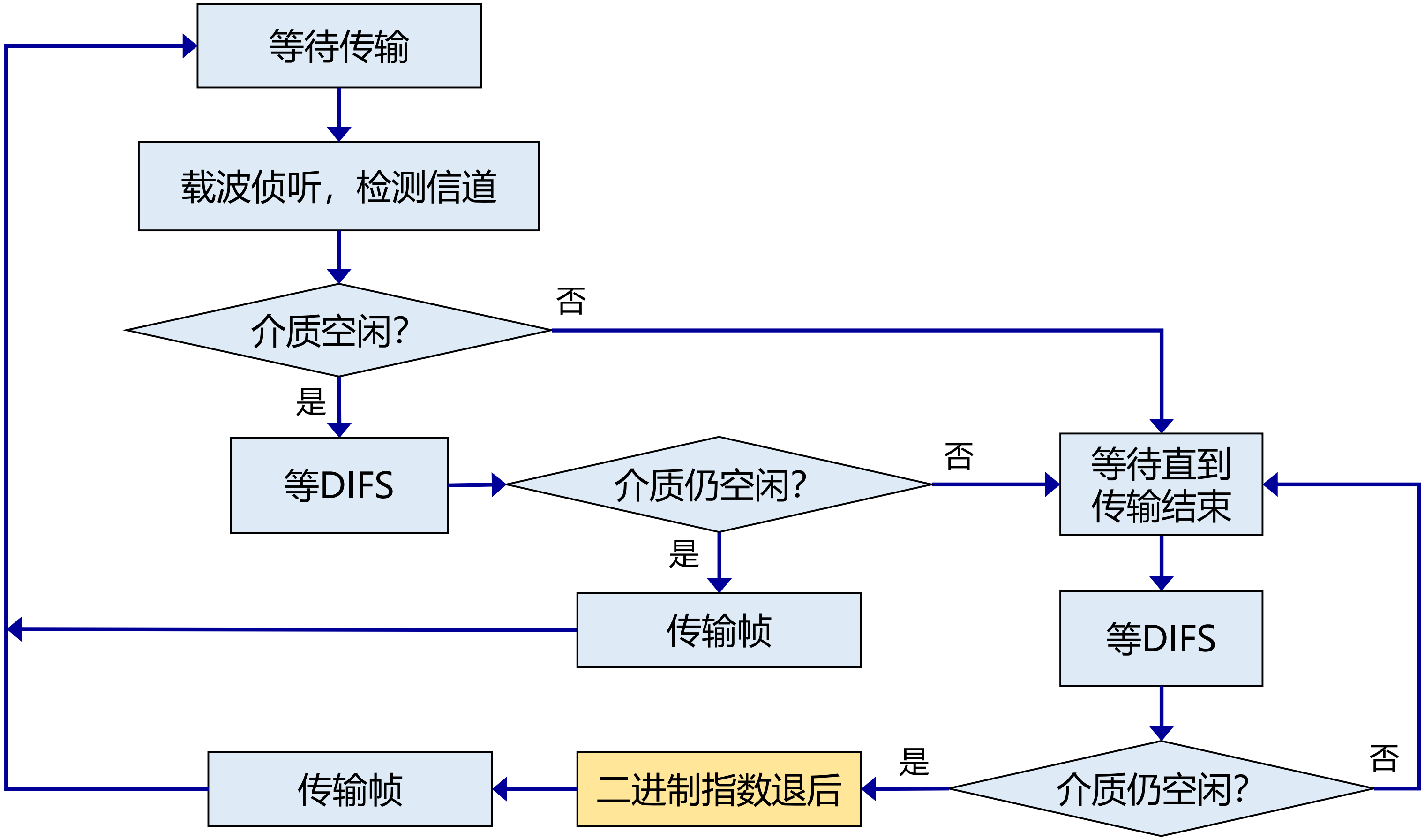
802.11 receiver

\- if frame received OK

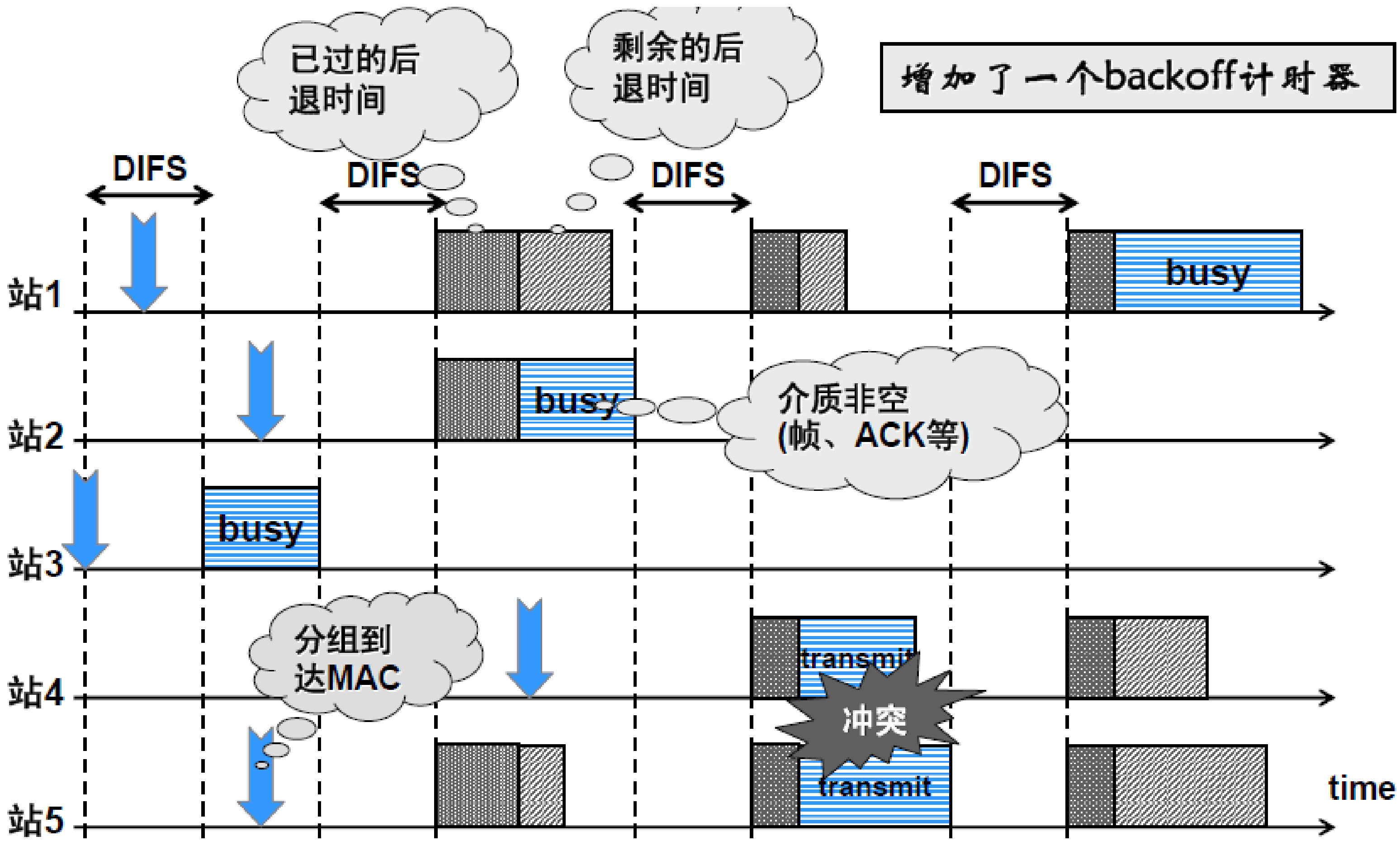
return ACK after SIFS (ACK needed due to hidden terminal problem)



介质访问控制方法: CSMA/CA



介质访问控制方法: CSMA/CA示例



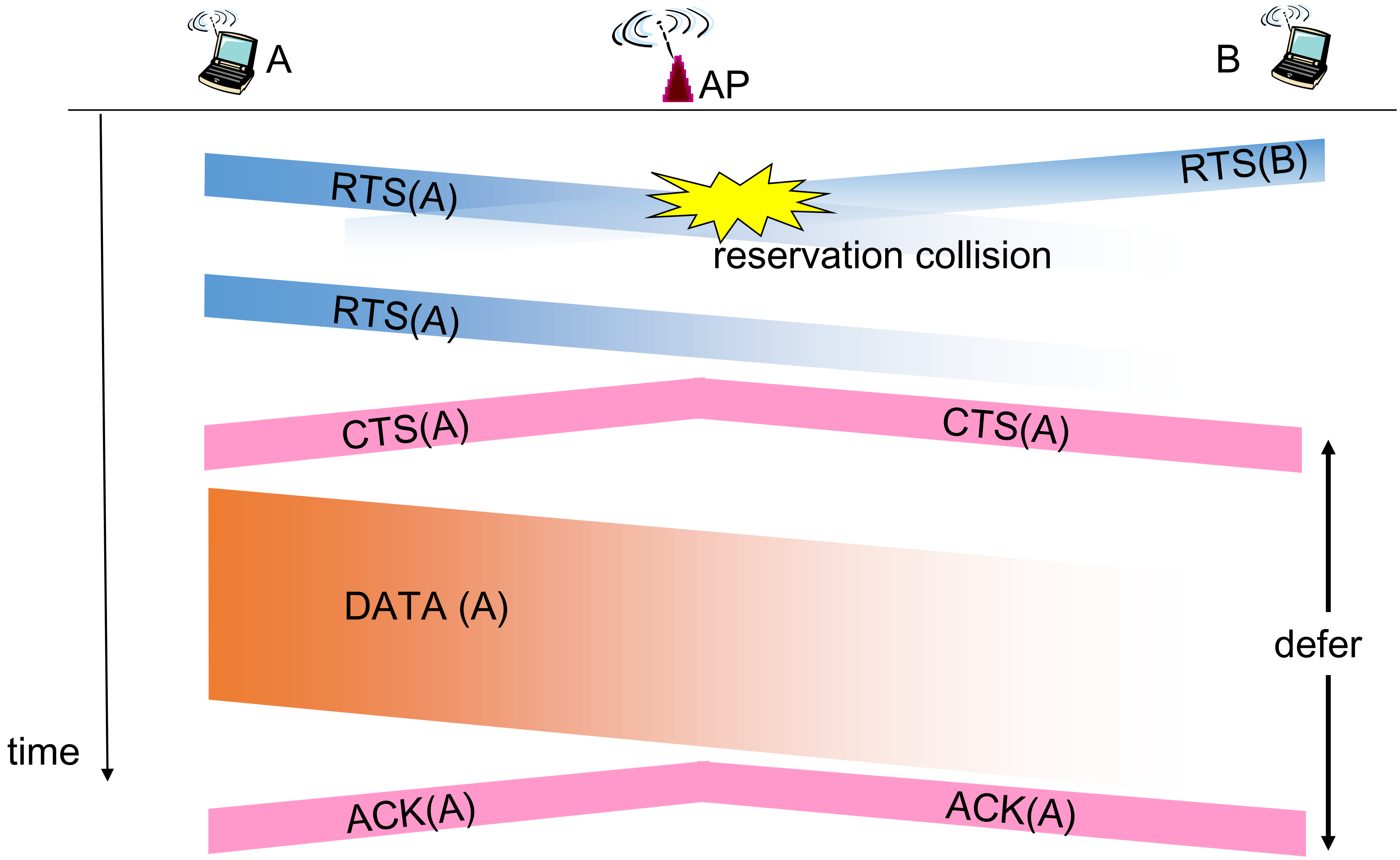
Avoiding collisions

idea: allow sender to "reserve" channel rather than random access of data frames: avoid collisions of long data frames

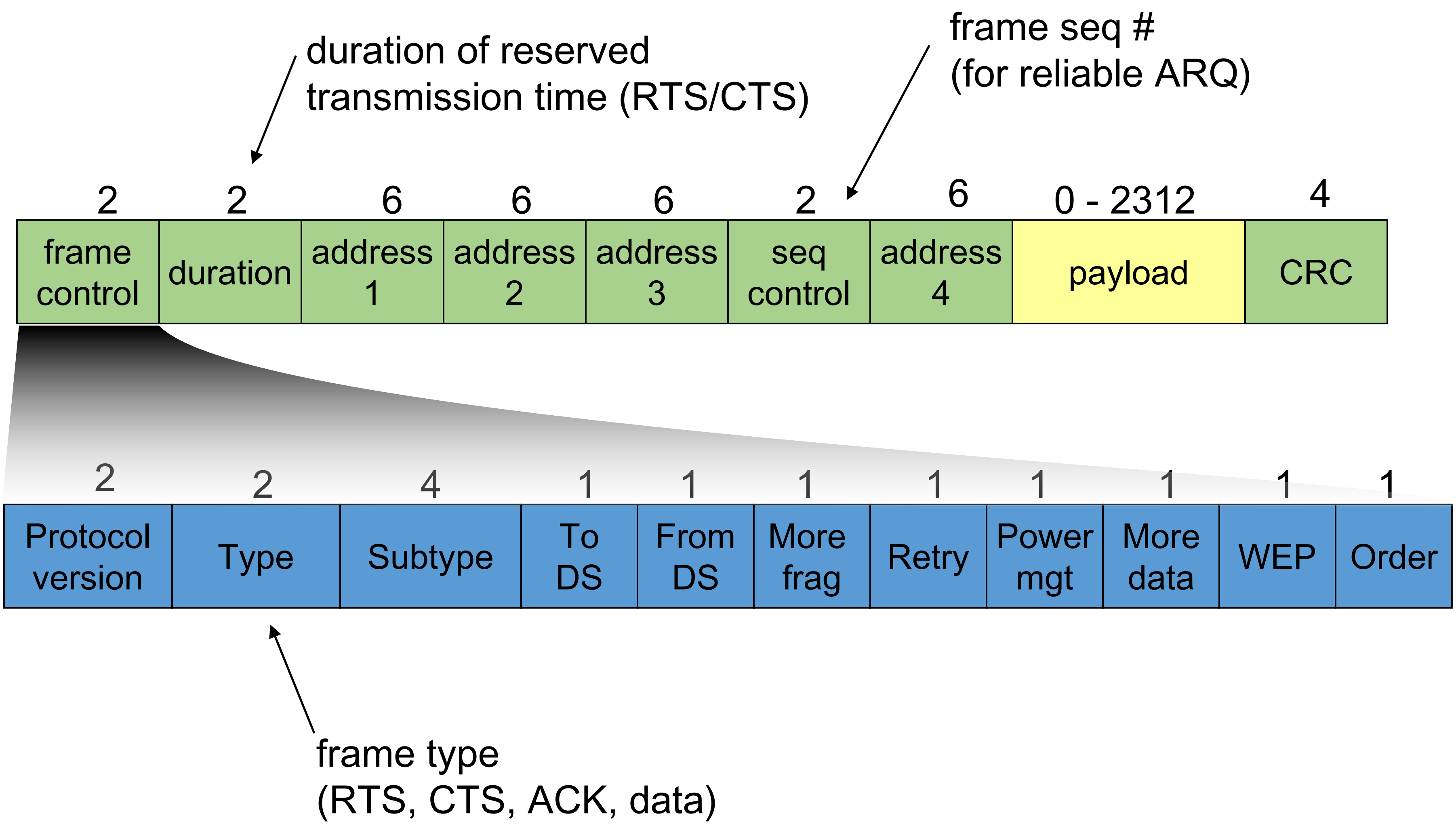
- sender first transmits *small* request-to-send (RTS) frame to AP using CSMA
  - RTSs may still collide with each other (but they're short)
- AP broadcasts clear-to-send CTS in response to RTS
- CTS heard by all nodes
  - sender transmits data frame
  - other stations defer transmissions

Avoid data frame collisions completely using small reservation frames!

Collision Avoidance: RTS-CTS exchange



802.11 frame


duration of reserved
transmission time (RTS/CTS)
frame seq #
(for reliable ARQ)
2
2
6
6
6
2
6
0 - 2312
4
frame control
duration
address 1
address 2
address 3
seq control
address 4
payload
CRC
2
2
4
1
1
1
1
1
1
1
1
Protocol version
Type
Subtype
To DS
From DS
More frag
Retry
Power mgt
More data
WEP
Order
frame type
(RTS, CTS, ACK, data)

802.11 frame: addressing

2	2	6	6	6	2	6	0 - 2312	4
frame control	duration	address 1	address 2	address 3	seq control	address 4	payload	CRC

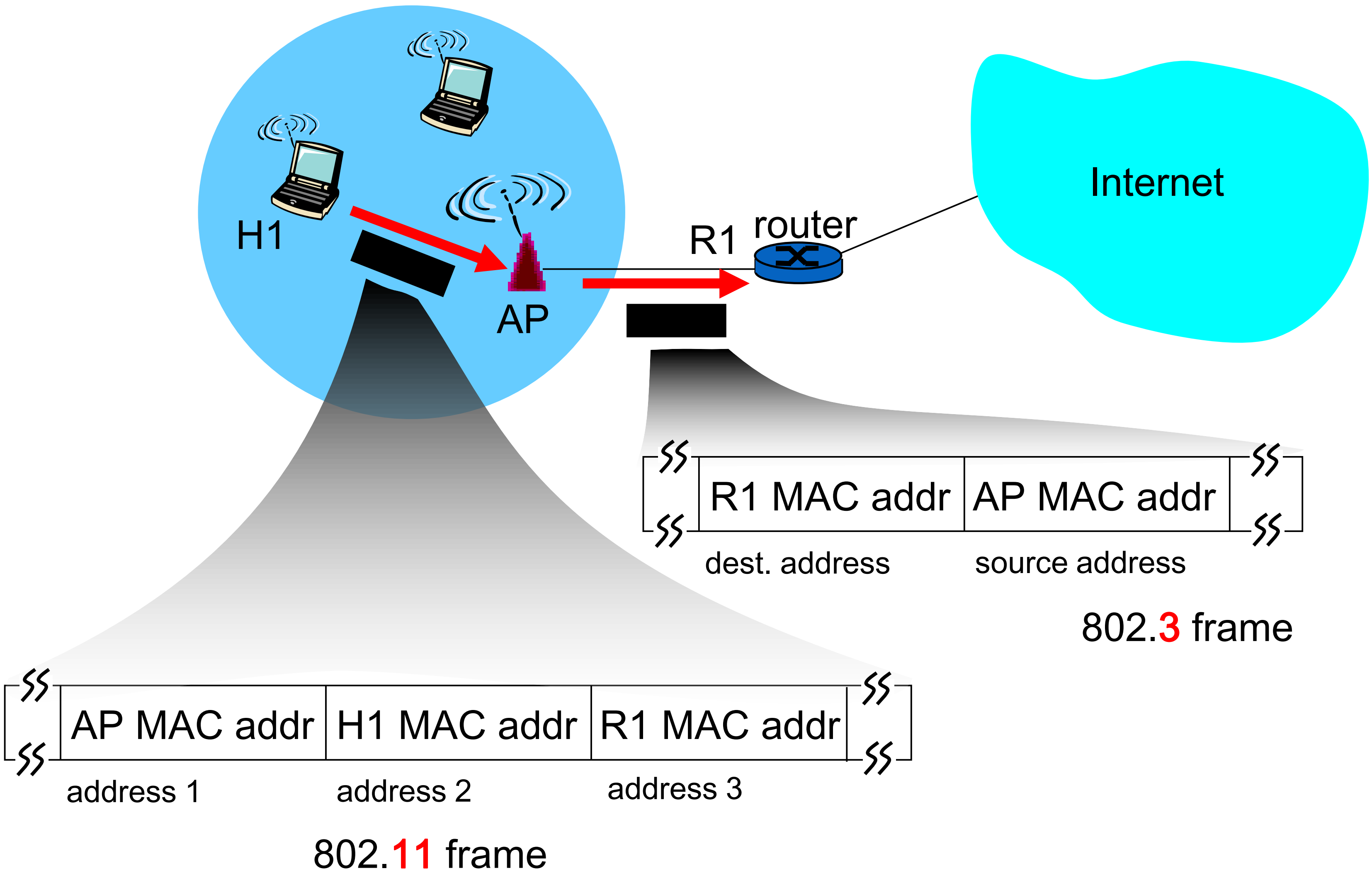
Address 1: MAC address of wireless host or AP to receive this frame

Address 2: MAC address of wireless host or AP transmitting this frame

Address 3: MAC address of router interface to which AP is attached

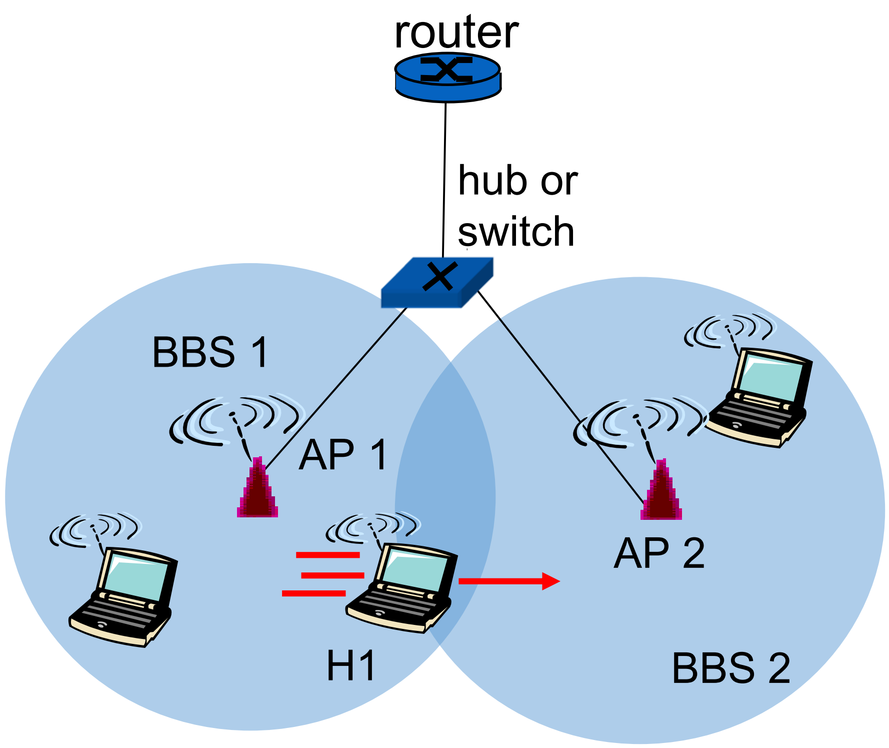
To DS	From DS	Address 1	Address 2	Address 3	Address 4
0	0	RA = DA	TA = SA	BSSID	N/A
0	1	RA = DA	TA = BSSID	SA	N/A
1	0	RA = BSSID	TA = SA	DA	N/A
1	1	RA	TA	DA	SA

802.11 frame: addressing



802.11: mobility within same subnet

- H1 remains in same IP subnet: IP address can remain same
- switch: which AP is associated with H1?
  - self-learning: switch will see frame from H1 and “remember” which switch port can be used to reach H1



总 结

- 接口层基础
- 局域网体系结构与组网方法
- 局域网编址与ARP协议
- 链路层差错控制
- 共享式以太网
- 交换式以太网
- 虚拟局域网
- 无线局域网