第5章 链路层：链路、接入网和局域网

学习目的

- **目的1：理解链路层服务的主要功能**
  - **差错检查、纠错**
  - **共享广播信道：多点接入问题(multiple access)**
  - **链路层寻址(link layer addressing)**
  - **局域网技术：Ethernet, VLANs**
- **目的2：链路层技术的实现**

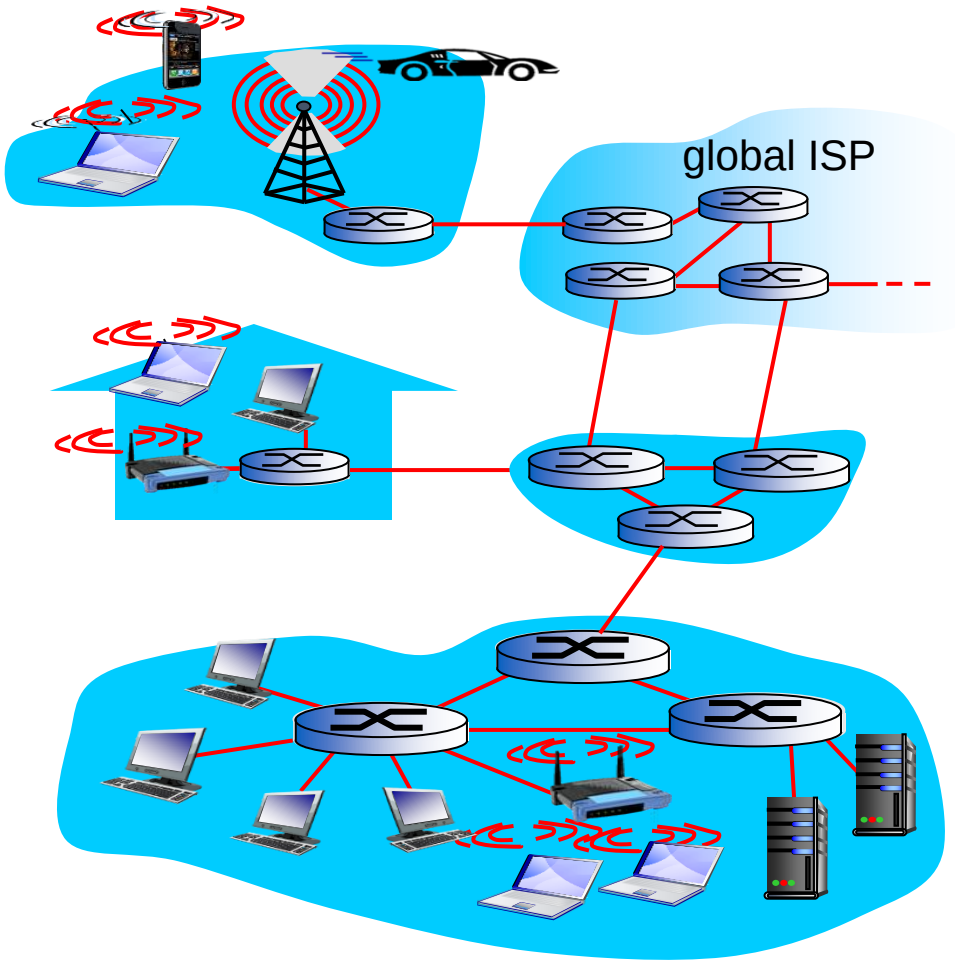
目录 CONTENT

5.1 链路层概述

链路层的术语

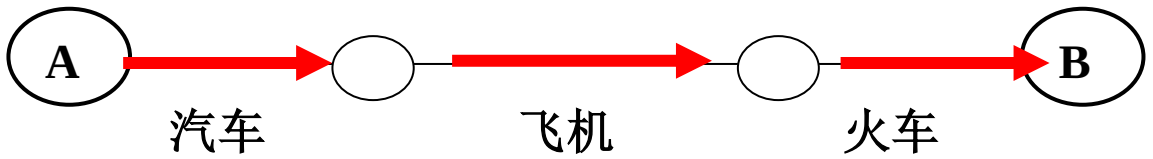
- 主机和路由器: **节点(nodes)**
- 沿着通信路径连接相邻节点的通信信道: **链路(links)**
  - 有线链路(wired links)
  - 无线链路(wireless links)
- 第二层的分组: **数据帧(frame)**, 它是封装了的数据报

数据链路层的职责是将数据报从一个节点传送到与该节点直接有物理链路相连的另一个节点。



链路层的类比

- 数据报可以在不同的链路上，通过不同的链路层协议发送：
  - e.g., Ethernet on first link, frame relay on intermediate links, 802.11 on last link
- 每个链路层协议提供不同的服务：
  - e.g., 可以提供/也可以不提供可靠数据传输服务



运输的类比:

- 从学校到洛桑的旅程
  - 小汽车: 电子科大——双流机场
  - 飞机: 双流机场——日内瓦
  - 火车: 日内瓦——洛桑
- 游客 = datagram
- 分段旅程 = communication link
- 运输模式 = link layer protocol
- 旅行社代理 = routing algorithm

5.1.1 链路层提供的服务

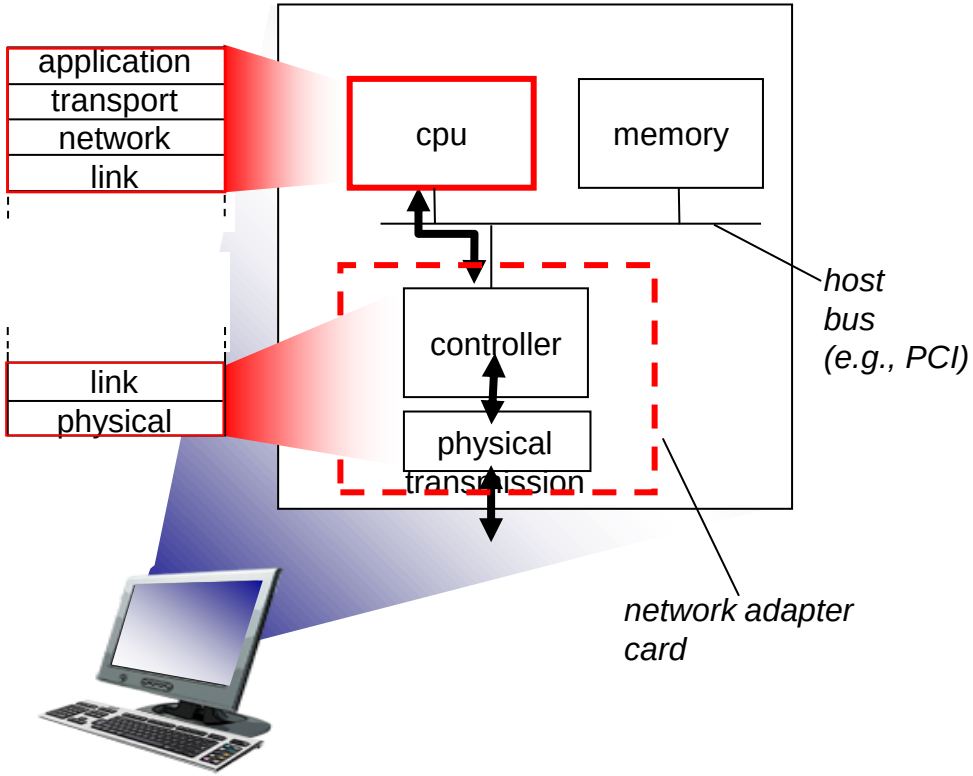
- **封装成帧，链路接入(framing, link access):**
  - 封装数据报为数据帧，增加头部，尾部信息
  - 如果是共享链路，接入链路
  - 在数据帧头部中，用MAC地址来标识源目的MAC地址
    - 不同于IP地址
- **在相邻节点之间可靠传输数据帧**
  - 我们在第3章已经学习了如何在运输层实现数据的可靠传输
  - 在比特错误率很低的链路(光纤、双绞线)很少使用
  - 无线链路：高比特错误率
    - 问题：为什么要在链路层和端到端都实现可靠传输?

5.1.1 链路层提供的服务

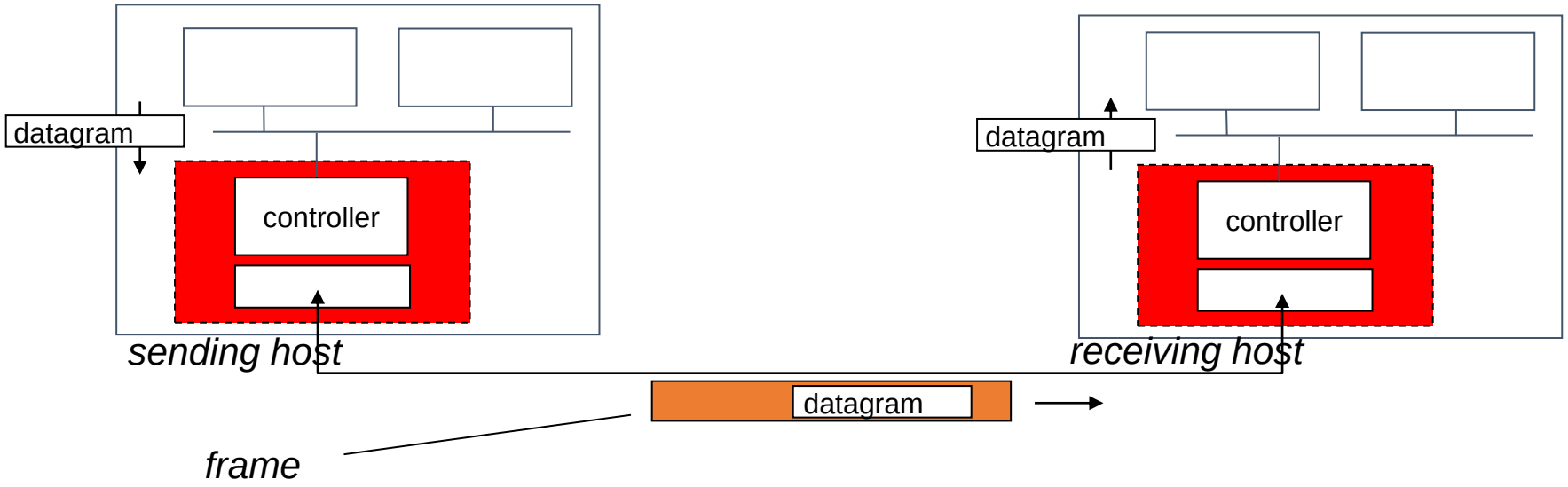
- **流量控制(flow control):**
  - 用于控制发送节点向直接相连的接收节点发送数据帧的频率
- **差错检查(error detection):**
  - 差错可能由信号衰减、噪声引入
  - 接收方检测是否出现错误：
    - 通知发送方重传或丢弃数据帧
- **错误纠正(error correction):**
  - 接收方标识和纠正比特错误，而不需要请求重传
- **半双工和全双工(half-duplex and full-duplex):**
  - 在半双工模式，链路的两个节点都可以发送数据，但是不能同时发送

5.1.2 链路层实现的位置

- 在主机和网络设备(路由器)上实现
- 在主机上，链路层的主体部分是在**网络适配器**上实现的(称为网卡)
  - 以太网卡，802.11卡；以太网芯片组
  - 实现链路层和物理层的功能
- 硬件、软件、固件的组合



网络适配器



发送方：

- 封装数据报为数据帧
- 增加差错检测比特，可靠数据传输，流量控制等机制。

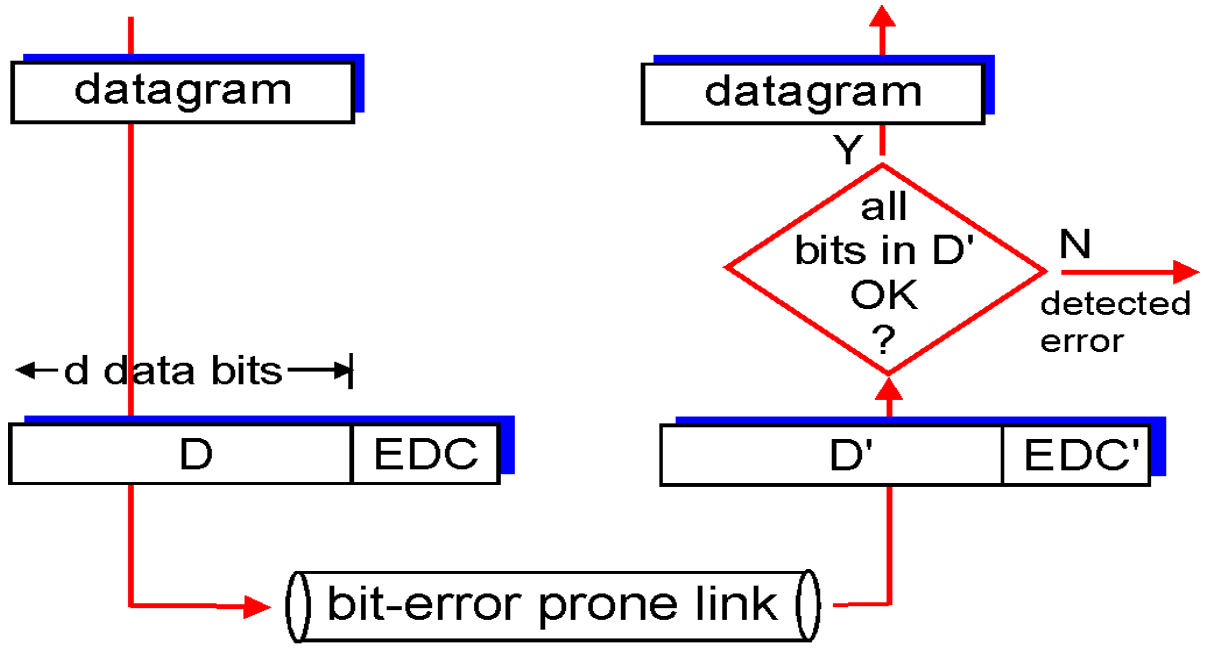
接收方

- 执行检查错误、可靠数据传输、流量控制等机制
- 抽取数据报，将其递交给上层

5.2 差错检测和纠错

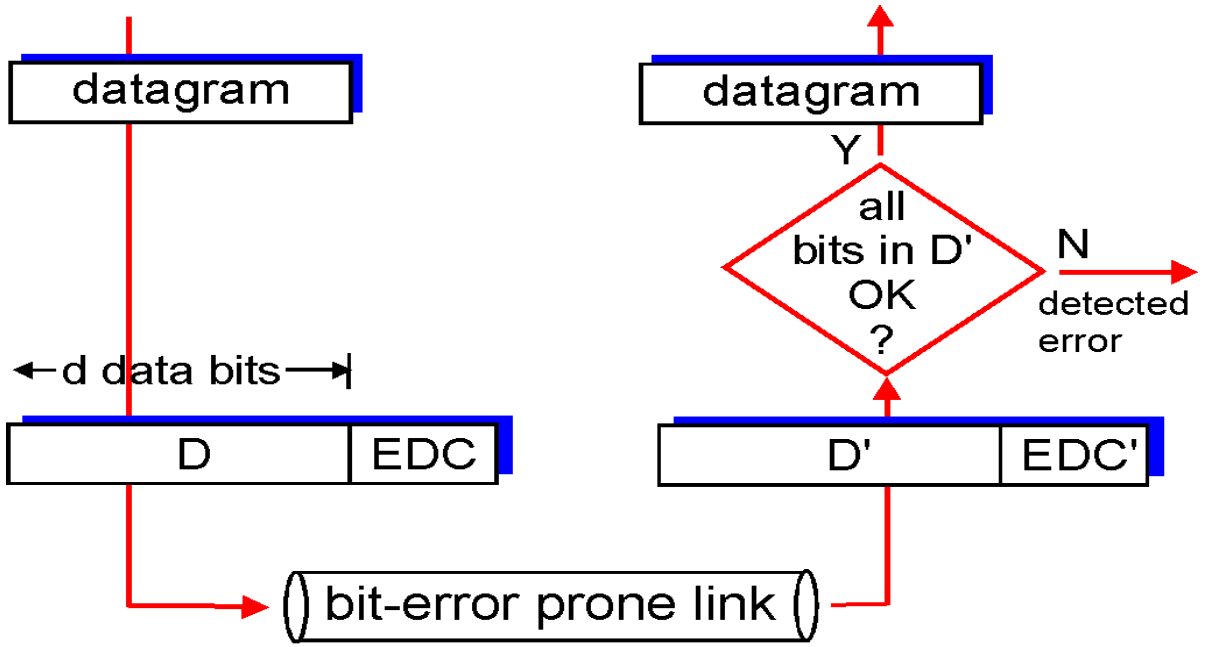
差错检测和纠错技术

- **比特级差错检测和纠错**
  - 对一个节点发送到一个相邻节点的帧，检测是否出现比特差错，并纠正。
  - 相关技术很多。
  - 差错检测和纠错的过程
- **差错检测并非100%可靠**
  - 协议可能丢失一些错误
  - 差错校验位越多，检测和纠正功能越好



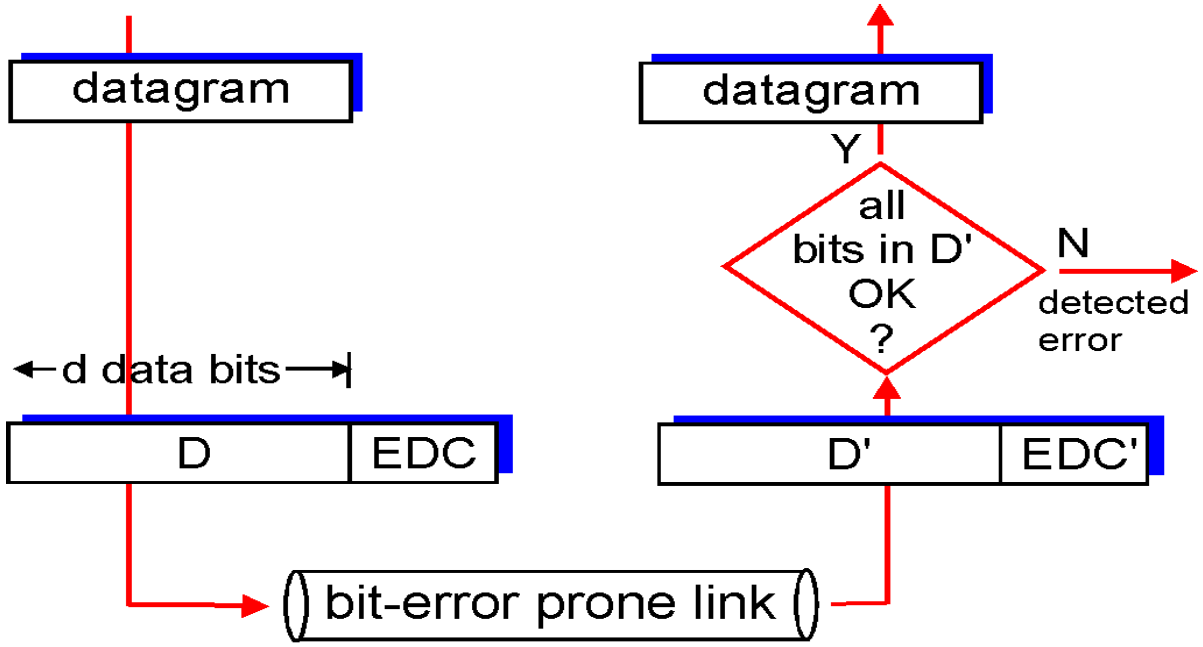
差错检测和纠错技术

- **发送节点**
  - 将数据D附加若干差错检测和纠错位EDC，一起发送到链路。
  - 数据D包括网络层传来的数据报，以及链路级寻址信息、序列号和其他字段。
  - 保护范围包括数据D的所有字段。



差错检测和纠错技术

- **接收节点**
  - 接收比特序列D'和EDC'。
  - 如果发生传输比特错误（0→1，1→0），D'和EDC'可能与发送的D和EDC不同。
  - 接收方根据D'和EDC'，判断D'是否和初始的D相同（D的传输是否正确）。
- 正确：解封取出数据报，交给网络层；
- 出错：差错处理。



说明

- 差错检测和纠正技术不能保证接收方检测到所有的比特差错，即*可能出现未检测到的比特差错*，而接收方并未发现。
- 选择一个合适的差错检测方案使未检测到的情况发生的概率很小即可。
- 差错检测和纠错技术越好，越复杂，开销更大。

三种主要差错检测技术

- **奇偶校验**：最基本的方法。
- **Internet校验和**：常用于运输层。
- **循环冗余检测**：常用于链路层。

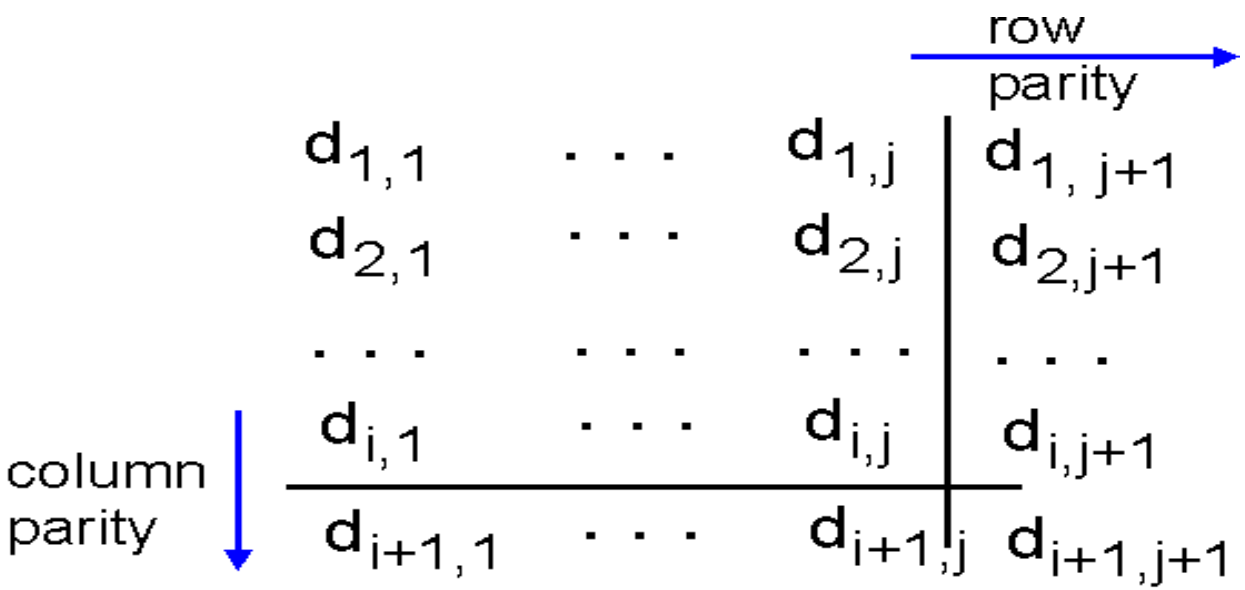
一比特奇偶校验

- **发送方：**
  - 在要发送的信息D（d位）后面附加一个奇偶校验位
  - 使“1”的个数是奇数（奇校验）或偶数（偶校验）
  - 一起传输发送（d+1位）。

- **接收方：**
  - 检测收到的信息（d+1位）中“1”的个数。
  - 偶校验：发现奇数个“1”，至少有一个比特发生差错（奇数个比特差错）。
  - 奇校验：发现偶数个“1”，至少有一个比特发生差错。
- 可以查出任意奇数个错误，但不能发现偶数个错误。
- 若比特差错概率很小，差错独立发生，一比特奇偶校验可满足要求。
- 若差错集中一起“突发”（突发差错），一帧中未检测到的差错的概率达到50%。

二维奇偶校验

- **基本思想：**
  - 将要传信息D（d比特）划分为i行j列（i 个组，每组j位）；
  - 对每行和每列分别计算奇偶值；
  - 结果的i+j+1个奇偶比特构成了帧的差错检测比特。



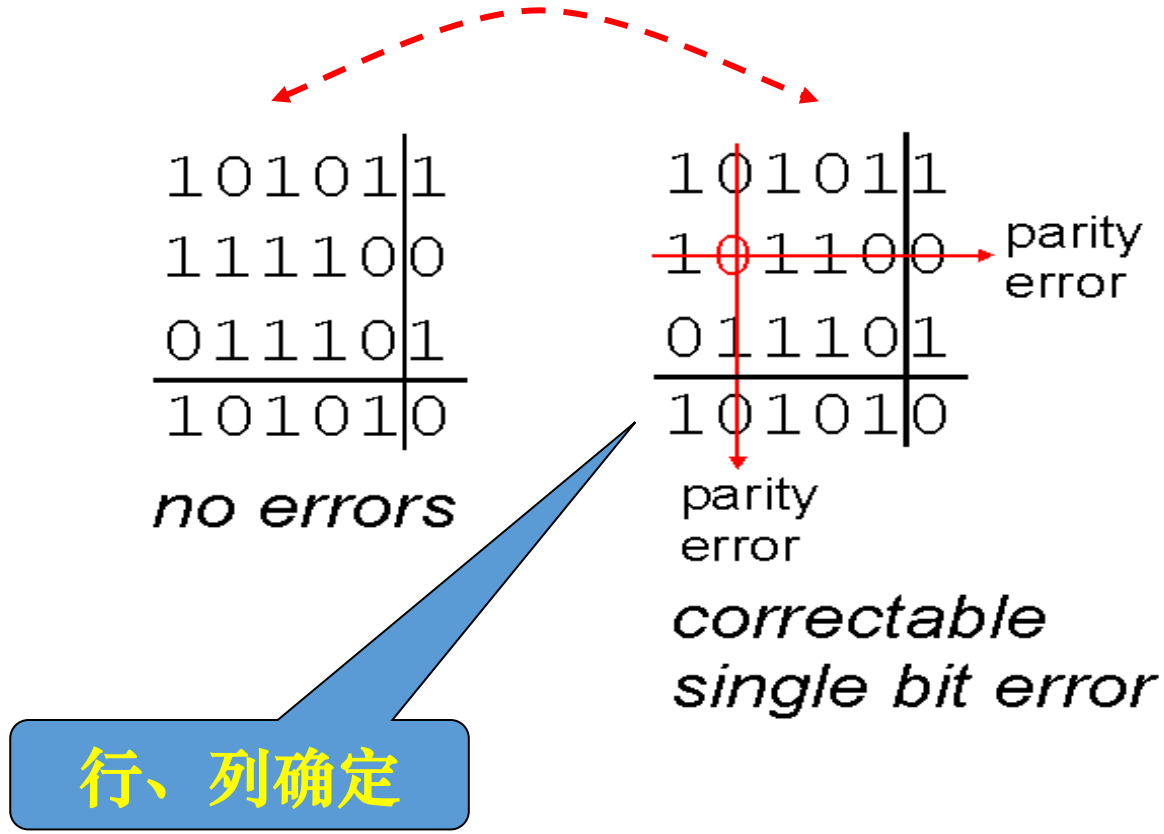
二维奇偶校验的例子

要发送的数据比特10101 11110 01110，

划分3组，每组5个比特。进行行、列偶校验

特点：

- 可以检测并纠正单个比特差错（数据或校验位中）。
- 能够检测(但不能纠正)分组中任意两个比特的差错。



Internet校验和方法（复习）

发送方：

- 将数据的每两个字节当作一个16位的整数，可分成若干整数；
- 将所有16 位的整数求和；
- 对得到的和逐位取反，作为检查和，放在报文段首部，一起发送。

接收方：

对接收到的信息 (包括检查和)按与发送方相同的方法求和。

- 全“1”：收到的数据无差错；
- 其中有“0”：收到的数据出现差错。

或者核对计算的检查和是否等于检查和字段的值。

Internet校验和的特点

- 分组开销小：检查和位数比较少；
- 差错检测能力弱：
- 适用于运输层（差错检测用软件实现，检查和方法简单、快速）。
- 链路层的差错检测由适配器中专用的硬件实现，采用更强的CRC方法。

循环冗余检测

$10111 \rightarrow x^4+x^2+x+1$

计算机网络中广泛采用

- **循环冗余检测CRC (cyclic redundancy check)编码：**
  - 即*多项式编码*，把要发送的比特串看作为系数是0或1的一个多项式，对比特串的操作看作为多项式运算。
- **基本思想：**
  - 设发送节点要把数据*D*（*d*比特）发送给接收节点。
  - 发送方和接收方先共同选定一个*生成多项式 G*（*r*+1比特），最高有效位 (最左边)是1。

循环冗余检测的基本思想

发送方：

- 计算出一个r位附加比特R，添加到D的后面产生DR（d+r 比特）
- DR能被生成多项式G模2运算整除，一起发送。

- **接收方**：
  - 用生成多项式G去除接收到的DR（d+r比特）
    - 余数非0：传输发生差错；
    - 余数为0：传输正确，去掉尾部r位，得所需数据D。

D：要发送的数据（d位）　R：CRC校验（r位）

DR（d+r位）

什么是模2运算

- 加法不进位，减法不借位，即操作数按位异或 (XOR)

例

1011 XOR 0101=1110 ； 1011-0101=1110

1001 XOR 1101=0100 ； 1001-1101=0100

- 乘法和除法与二进制运算类似，其中加法或减法没有进位或借位。
- 乘以2^r，即比特模式左移r个位置。

$$D\times 2^r \text{ XOR } R = D\,00\ldots00 \text{ XOR } R$$

$$= DR \ (d+r\text{比特})$$

计算R（CRC比特）

- DR能被G模2运算整除：即

$$D \times 2^r \ \text{XOR} \ R = nG$$

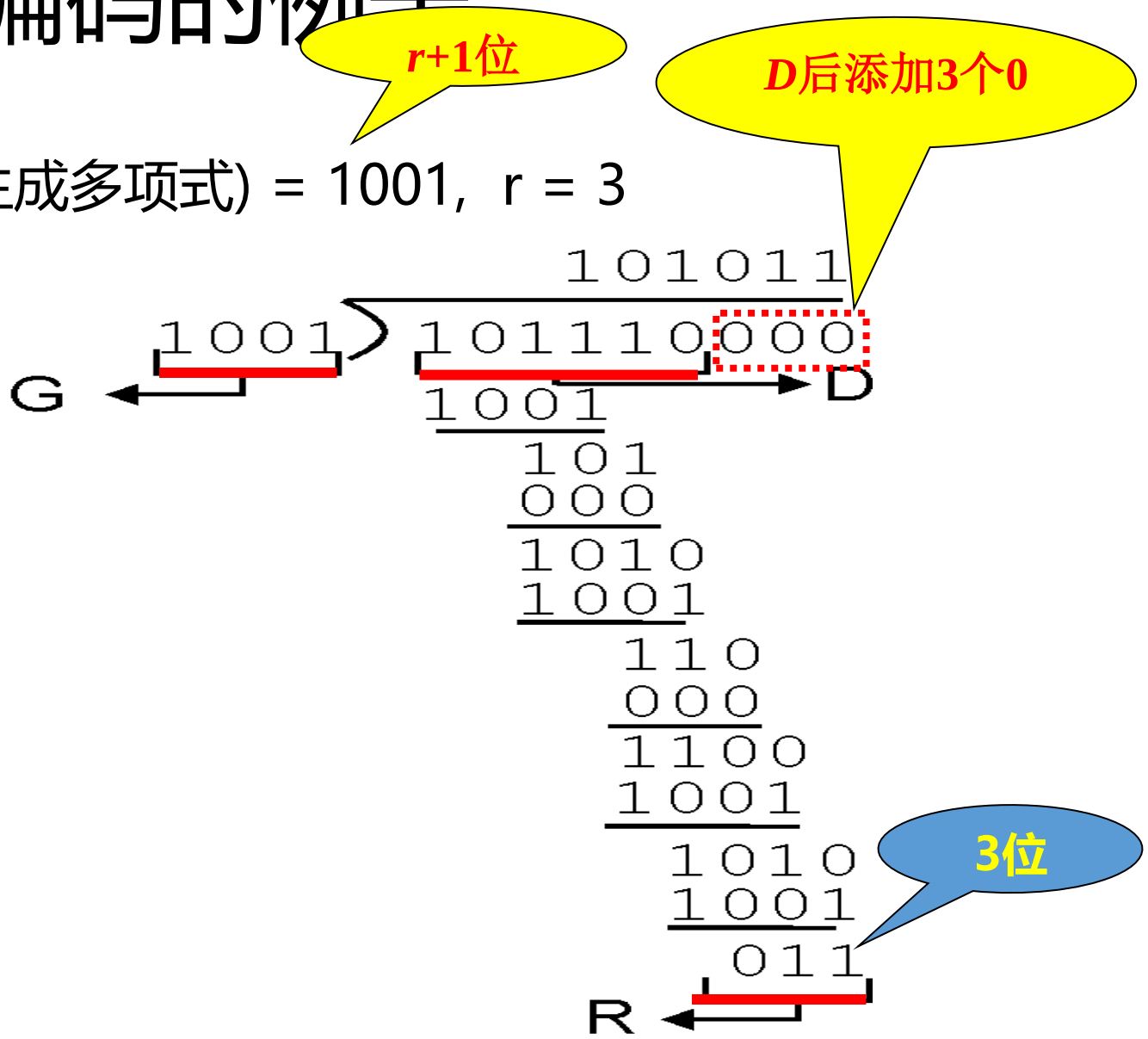
- 等式两边都用R异或，得到

$$D \times 2^r = nG \ \text{XOR} \ R$$

- 即用G来除$D \times 2^r$，余数值刚好为R。
- **R的计算：将数据D后面添加r个0，除以给定的生成多项式G，所得余数即为R（r位）。**

CRC编码的例子

设 (数据)　D = 101110,d = 6,　G (生成多项式) = 1001,　r = 3



实际传输的数据形式是:
101110 011

CRC练习

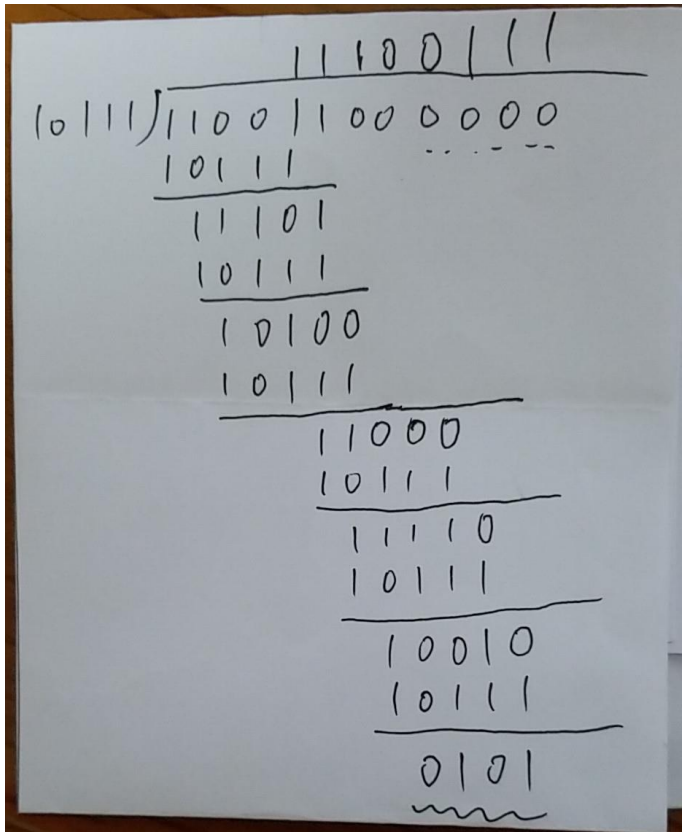
假设:

- 通信双方协商的生成多项式为:
  $G=X^4+X^2+X+1$
- 发送方要发送的数据为:
  D=11001100

问题:

- CRC校验信息需要多少位?
- 发送方最终发送的数据是多少位?
- 发送方最终发送的数据内容是什么?
- 如果传输过程中出现一位错误能否检测?出现六位错误能否检测?

CRC练习



假设:

- 通信双方协商的生成多项式为:
  $G=X^4+X^2+X+1$
- 发送方要发送的数据为:
  D=11001100

问题答案:

- CRC校验信息需要多少位?—— 4位
- 发送方最终发送的数据是多少位?—— 12位
- 发送方最终发送的数据内容是什么?——11001100 0101
- 如果传输过程中出现一位错误能否检测?出现六位错误能否检测?

—— 一位错误可检测到，六位错误一般情况下不可以（见教材P292）

循环冗余码CRC的特点

- 生成多项式G的选择：常见的有8、12、16和32 比特生成多项式G。
- 国际标准已经定义了8-、16-、32-位生成多项式G；8-位CRC用于ATM信元首部的保护；32-CRC用于大量链路层IEEE协议。其他检错方法不常用，故不作专门介绍
  - CRC8生成多项式为G(x)=x8+x5+x4+1
  - CRC16生成多项式为G(x)=x16+x12+x5+1
  - CRC-32生成多项式为
    G(x)=x32+x26+x23+x22+x16+x11+x10+x16+x8+x7+x5+x4+x2+x+1
- CRC特点：

  能检测小于 r+1 位的突发差错、任何奇数个差错。

差错检测方法比较

- 奇偶校验能力最弱，CRC校验能力最强。
- 奇偶校验通常用于简单的串口通信
- Internet校验和通常用于网络层及其之上的层次，要求简单快速的软件实现方式
- CRC通常应用于链路层，一般由适配器硬件实现

检错纠错的基本原理*

- 为了校错和纠错，我们在发送方对信息位进行编码，而在接收方进行解码，以还原信息。编码的原理是增加冗余信息检验位。
- 假设要传送的信息为m位，校验位为r位，则编码的总长度为m + r位。所谓编码就是把m位信息码映射为m + r位的编码。
- m位信息仅有2^m种信息码，而m + r位则有2^{m+r}种编码。因此把m + r位的编码分为两大类：有效码和无效码，有效码指与信息码有一一映射关系的编码，所以有效码有2^m个。而其余的编码则称之为无效码，无效码共有$2^n - 2^m = 2^{m+r} - 2^m = (2^r - 1)2^m$个。
- 发方进行编码就是把信息码映射为有效码。

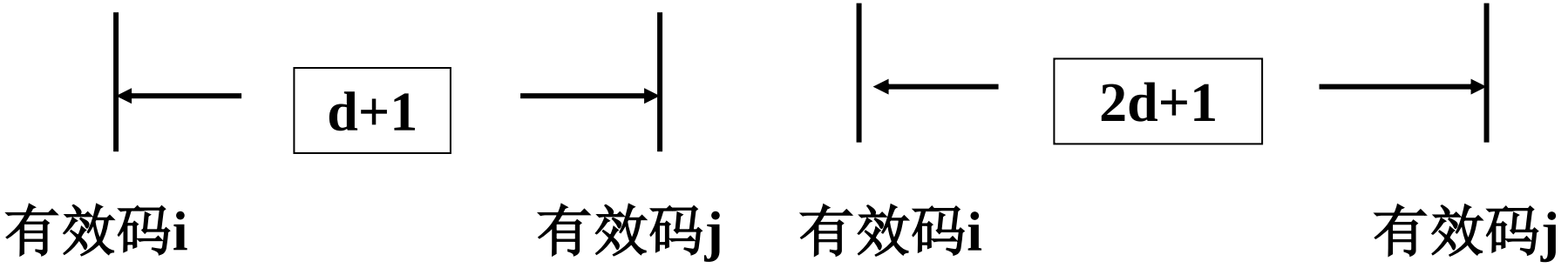
检错纠错的基本原理*

- 由于信道有错，有效码在传送过程中有可能变成了无效码
- 当接收方收到信道传来的编码信息时，首先判断它是有效码还是无效码。
  - 如果是有效码则通过逆映射关系，解码出原信息
  - 如果是无效码，则认为出了错，这就是检验出错误
- 假设通过无效码能够分析出它是由哪一个有效码出错而来的，从而找出原信息码，这就称为纠错（编码）
- 如果信道出错使一个有效码变成另一个有效码，则收方检不出错。

海明编码*

- 我们把两个等长二进制数（或两个等长码）不相同对应位的位数称为距离。例如10000和00100的距离为2。
- [定义] 海明距离(Hamming Distance)：

  某种编码任意两个有效码间的最小距离称为该编码的海明距离。
- 结论1：可以检出d个错误的检错码，其海明距离至少为d + 1
- 结论2：可以纠出d个错误的纠错码，其海明距离至少为2d + 1
- 海明码：能够纠正一位错误的编码称为海明码。



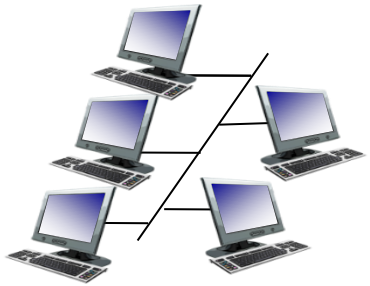
5.3 多路访问链路和协议

多路访问链路和协议

- 两种网络链路：
  - 点对点链路：链路两端各一个节点。一个发送和一个接收。如点对点协议PPP。
  - 广播链路： 多个节点连接到一个共享的广播信道。

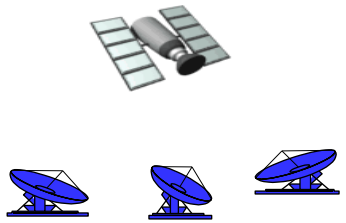
广播：任何一个节点传输一帧时，信号在信道上广播，其他节点都可以收到一个拷贝。常用于局域网LAN中，如早期的以太网和无线局域网。

本节主要学习**广播链路的信道共享技术**。



shared wire (e.g., cabled Ethernet)

shared RF (e.g., 802.11 WiFi)



shared RF (satellite)

humans at a cocktail party (shared air, acoustical)

广播信道要解决问题

- **传统的广播电视**：是单向的广播，一个固定的节点向许多接收节点发送。
- **计算机网络**：广播信道上的节点都能够发送和接收。
  - 好比许多人聚集在一起交谈（空气是广播媒体）。
  - 解决“谁在什么时候获得说话权力”(向信道发送)。
- **多路访问问题**：如何协调多个发送和接收节点对共享广播信道的访问。相关技术即是多路访问协议（也称多址访问协议）。

多路访问协议

- 目的：协调多个节点在共享广播信道上的传输。
  - 避免多个节点同时使用信道，发生冲突（碰撞），产生互相干扰。
- 冲突（collide）：两个以上的节点同时传输帧，使接收方收不到正确的帧（所有冲突的帧都受损丢失）。
  - 造成广播信道时间的浪费。
  - 多路访问协议可用于许多不同的网络环境，如有线和无线局域网、卫星网等。

多路访问协议

- 理想的多址访问协议

速率为R bps的广播信道

1. 当一个节点有数据发送时，它能以R bps的速率发送.
2. 当有M个节点要发送数据，每个节点的平均发送速率为 R/M
3. 完全分散:
   - 不需要主节点协调传输
   - 不需要时钟、时隙同步
4. 简单

多路访问协议类型（三类）

- 信道划分协议
  - 把信道划分为小“片” (时隙)
  - 给节点分配专用的小“片”
- 随机访问协议
  - 不划分信道，允许冲突
  - 能从冲突中“恢复”
- 轮流协议
  - 通过轮流访问信道避免冲突，要发送的节点越多轮流时间越长

5.3.1 信道划分协议

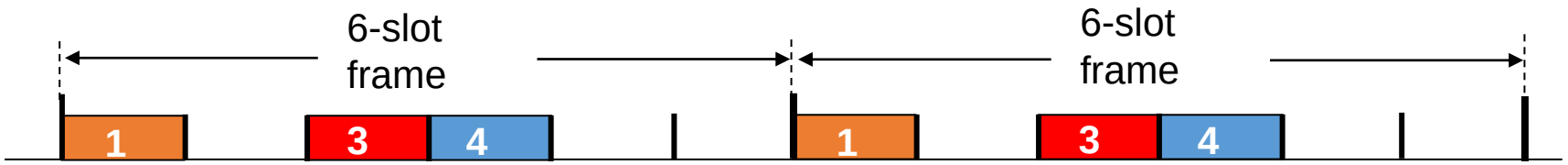
主要有TDMA、FDMA、CDMA三种。

设信道支持 *N* 个节点，传输速率是 *R* b/s。

- **时分多路访问TDMA (time division multiple access)：**

  将时间划分为*时间帧*，每个时间帧再划分为*N个时隙*（长度保证发送一个分组），分别分配给*N*个节点。每个节点只在固定分配的时隙中传输。

  例：6个站点的LAN, 时隙1、3、4 有分组, 时隙2、5、6 空闲



5.3.1 信道划分协议

TDMA的特点

- 避免冲突、公平：每个节点专用速率R/N b/s。
- 节点速率有限：R/N b/s；
- 效率不高：节点必须等待它的传输时隙。

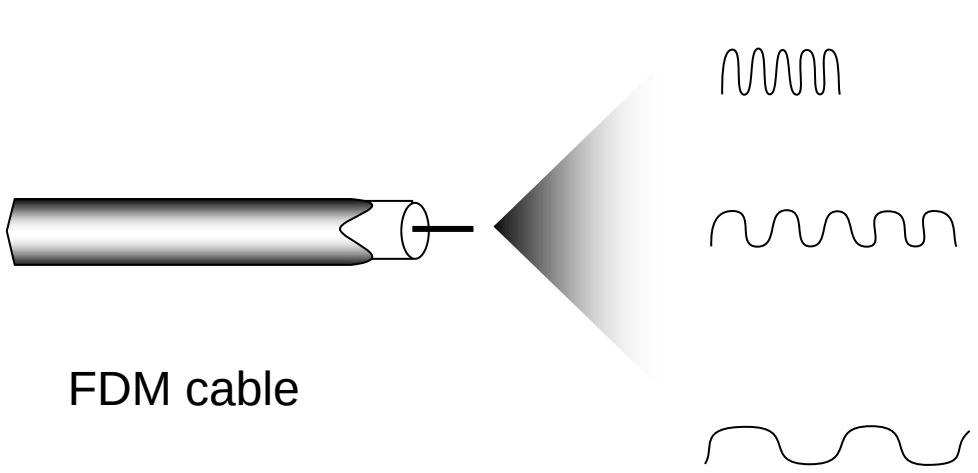
5.3.1 信道划分协议

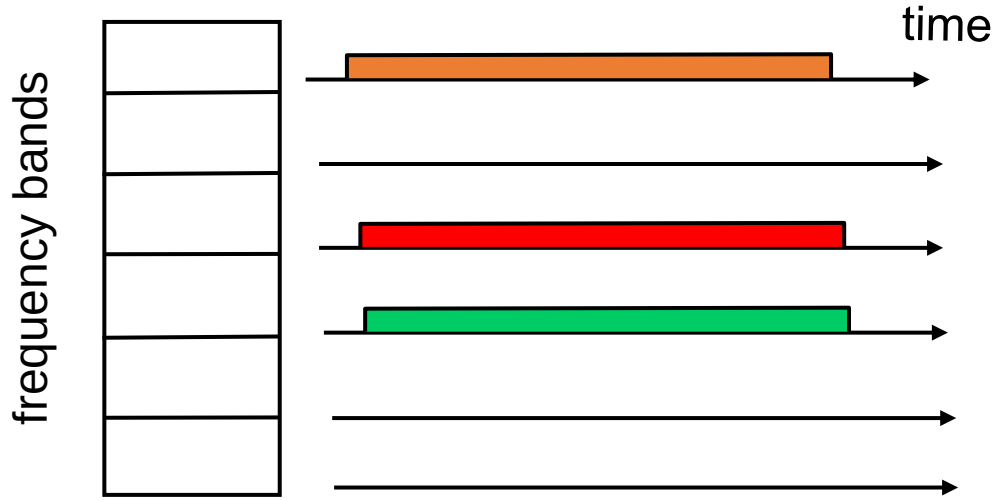
- **频分多路访问FDMA (frequency division multiple access)：**

将总信道带宽 R b/s划分为 N 个较小信道（频段，带宽为R/N），分别分配给 N 个节点。例： 6个站点的LAN, 频带1、3、4 有分组, 频带2、5、6 空闲

特点：

- 避免冲突、公平：N个节点公平划分带宽；
- 节点带宽有限、效率不高：节点带宽为R/N。





5.3.1 信道划分协议

- **码分多路访问CDMA (frequency division multiple access)：**
  - 每个节点分配一个唯一的编码
  - 每个节点用它唯一的编码来对它发送的数据进行编码
  - 允许多个节点“共存”，信号可叠加，即可以同时传输数据而无冲突(如果编码 是“正交化”的)

5.3.2 随机访问协议

基本思想：

- 发送节点以信道全部速率（R b/s）发送；
- 发生冲突时，冲突的每个节点分别等待一个随机时间，再重发，直到帧(分组)发送成功
- 节点间没有协调者

典型随机访问协议：

- ALOHA协议(纯ALOHA，时隙ALOHA)
- 载波监听多路访问CSMA协议
- 带冲突检测的载波监听多路访问CSMA/CD
- 带冲突避免的载波监听多路访问CSMA/CA

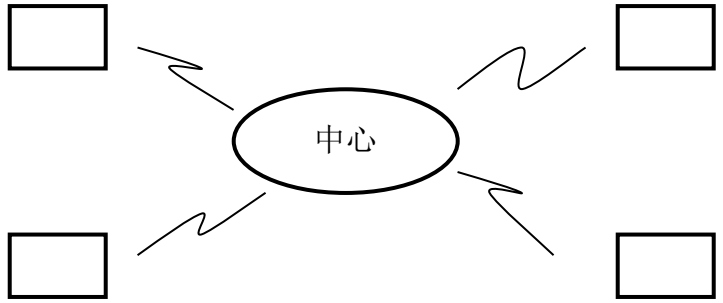
ALOHA

ALOHA：夏威夷大学研制的一个无线电广播通信网（20世纪70年代初），采用星型拓扑结构，使地理上分散的用户通过无线电来使用中心主机。

- 中心主机通过下行信道向二级主机广播分组；
- 二级主机通过上行信道向中心主机发送分组（可能会冲突，无线电信道是一个公用信道）。

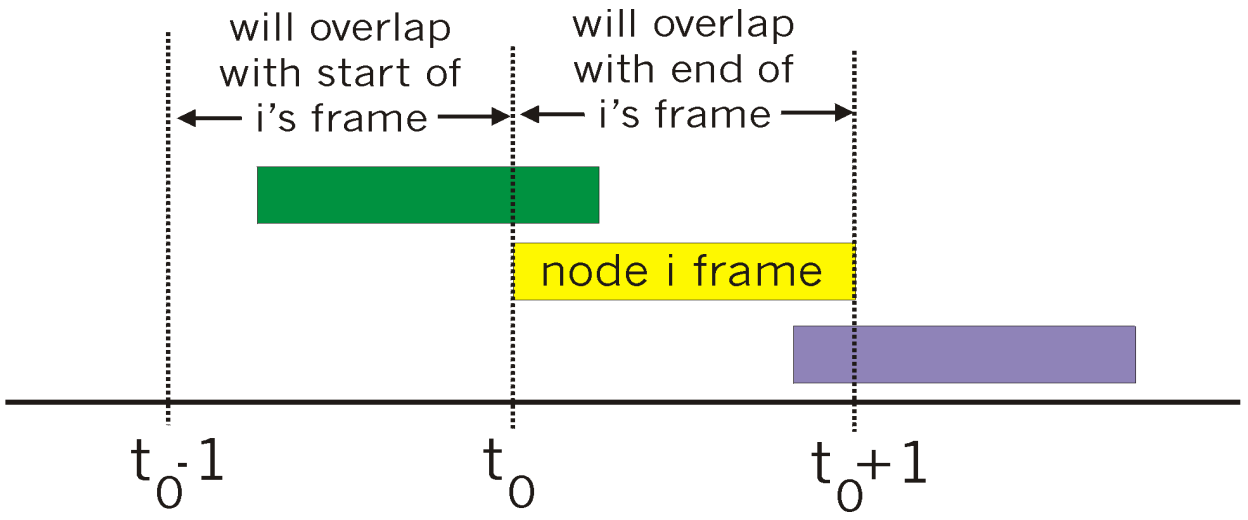
有两种形式：

- 纯ALOHA
- 时隙ALOHA



纯 ALOHA

- 非时隙Aloha: 简单，不需同步
- 帧一到达，立即传输
- 如果与其他帧产生冲突，在该冲突帧传完之后：
  - 以概率p立即重传该帧；
  - 或等待一个帧的传输时间，再以概率p传输该帧，或者以概率$1-p$等待另一个帧的时间。
- 冲突概率:
  - 在t_0发送的帧，和在 $[t_0-1,t_0+1]$ 的发送的其它帧冲突



纯Aloha效率

P(给定节点成功传送) = P(节点传送) x

P(没有其他节点在$[t_0-1,t_0]$内传送) x

P(没有其他节点在$[t_0,t_0+1]$内传送)

$= p \cdot (1-p)^{N-1} \cdot (1-p)^{N-1}$

$= p \cdot (1-p)^{2(N-1)}$

... 选择p值，然后求N->无穷时的极限...

= 1/(2e) = 0.18

纯Aloha效率

推导过程：

$$E(p) = Np(1-p)^{2(N-1)}$$

$$E'(p) = N(1-p)^{2(N-2)} - Np2(N-1)(1-p)^{2(N-3)}$$

$$= N(1-p)^{2(N-3)}((1-p) - p2(N-1))$$

$$E'(p) = 0 \Rightarrow p^* = \frac{1}{2N-1}$$

$$E(p^*) = \frac{N}{2N-1}(1 - \frac{1}{2N-1})^{2(N-1)}$$

$$\lim_{N\to\infty} E(p^*) = \frac{1}{2} \cdot \frac{1}{e} = \frac{1}{2e}$$

（本章课后习题第9题）

时隙ALOHA

假设

- 所有帧大小相同
- 时间被划分为相同大小的时隙，一个时隙等于传送一帧的时间
- 节点只能在一个时隙的开始才能传送
- 节点需要同步
- 如果一个时隙有多个节点同时传送，所有节点都能检测到冲突

实现

- 当节点要发送新帧，它等到下一时隙开始时传送
- 没有冲突，节点可以在下一时隙发送新帧
- 如果有冲突，节点在随后的时隙以概率p重传该帧，直到成功为止。

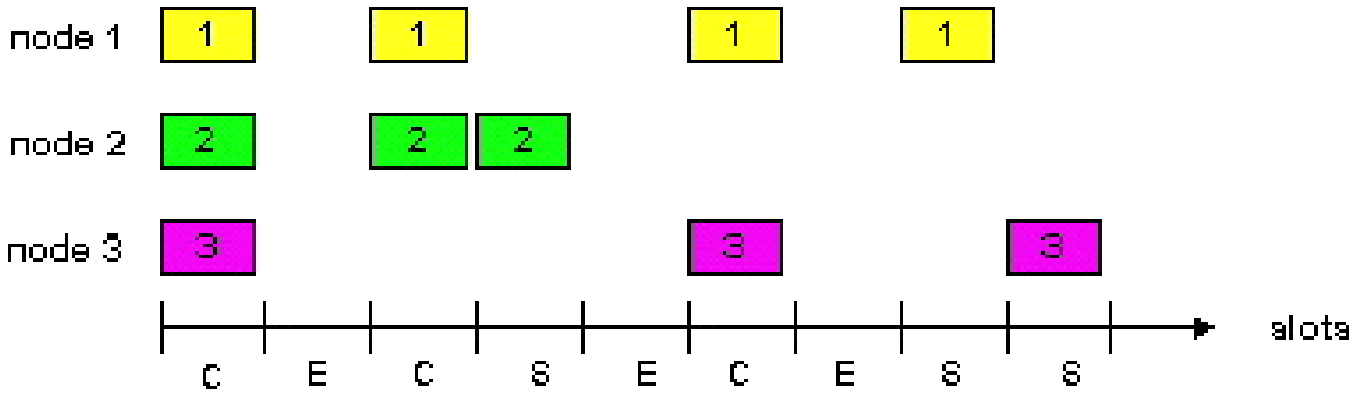
时隙ALOHA

优点

- 单个活跃节点可以持续以满速率传送帧
- 具有高分散性: 只需节点的时隙同步
- 简单

缺点

- 冲突，浪费时隙
- 空闲时隙
- 节点只有在传输数据包时才能检测到冲突



时隙Aloha效率

效率：当有很多节点，每个节点有很多帧要发送时，成功时隙所占的百分比

- 假设有N个节点，每个节点在时隙以概率*p*发送
- 一个节点在一个时隙成功传送的概率 = $p(1-p)^{N-1}$
- 任一节点传送成功的概率 = $Np(1-p)^{N-1}$

- 为了得到N个活跃节点的最大效率，必须找出使表达式$Np(1-p)^{N-1}$ 取最大值的p^*
- 为了得到大量活跃节点的最大效率， 我们求N趋近无穷时$Np^*(1-p^*)^{N-1}$ 极限值，计算可知最大效率为1/e =0 .37

最佳情况: 信道有 37% 的有效传输

课后练习*

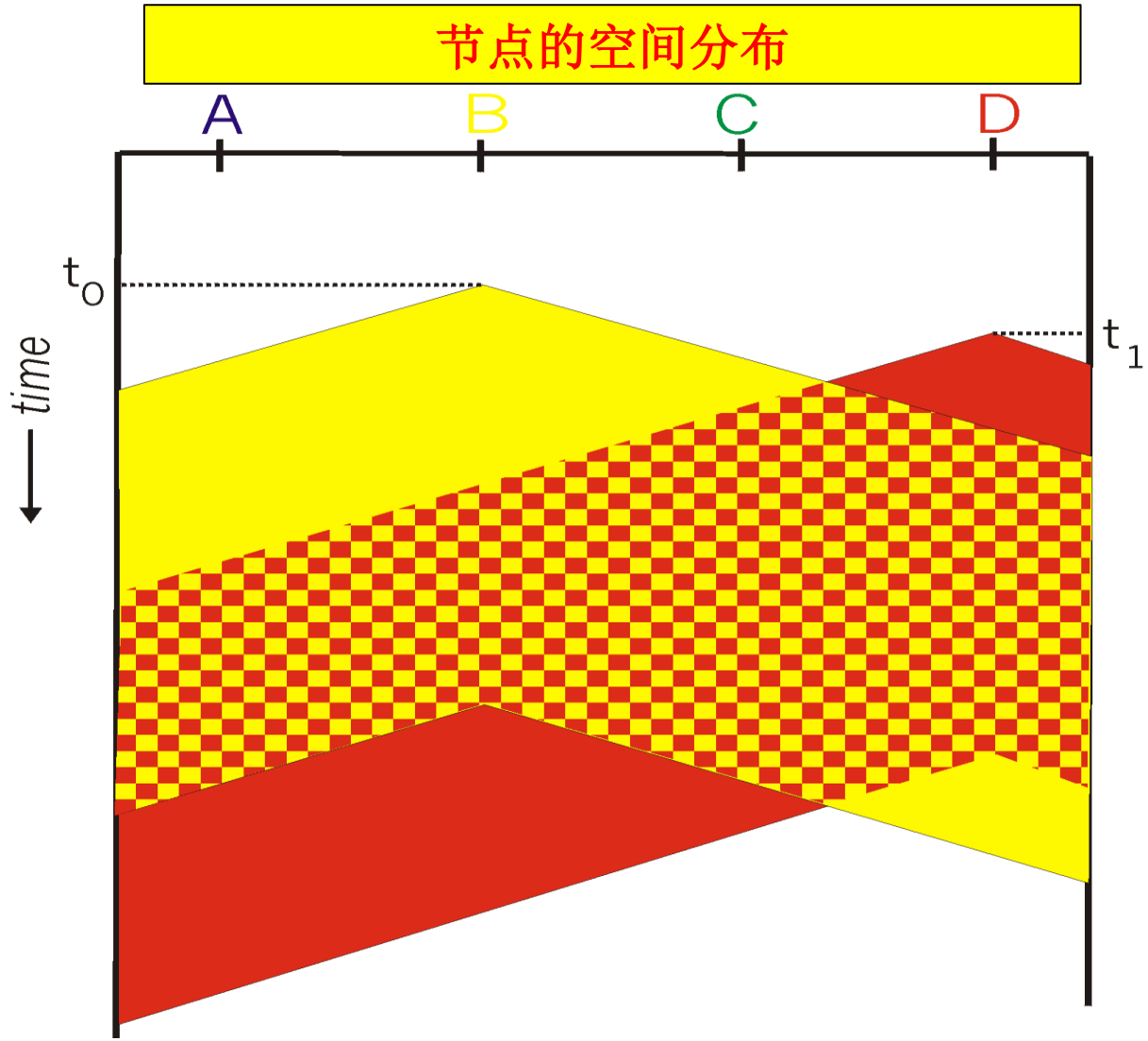
假设信道帧到达率服从泊松分布，试从单位时间内帧到达数和信道吞吐量的角度，推导纯Aloha协议及时隙Aloha协议的效率。

CSMA（载波侦听多路访问）

- **载波侦听**：某个节点在发送之前，先监听信道。
  - 信道忙：有其他节点正往信道发送帧，该节点随机等待（回退）一段时间，然后再侦听信道。
  - 信道空：该节点开始传输整个数据帧。
- **人类类比: 自己说话之前，先听一下有没有其他人正在说话，不要打断他人说话!**
- **CSMA 的特点:**
  - 发前监听，可减少冲突。
  - 由于传播时延的存在，仍有可能出现冲突，并造成信道浪费。

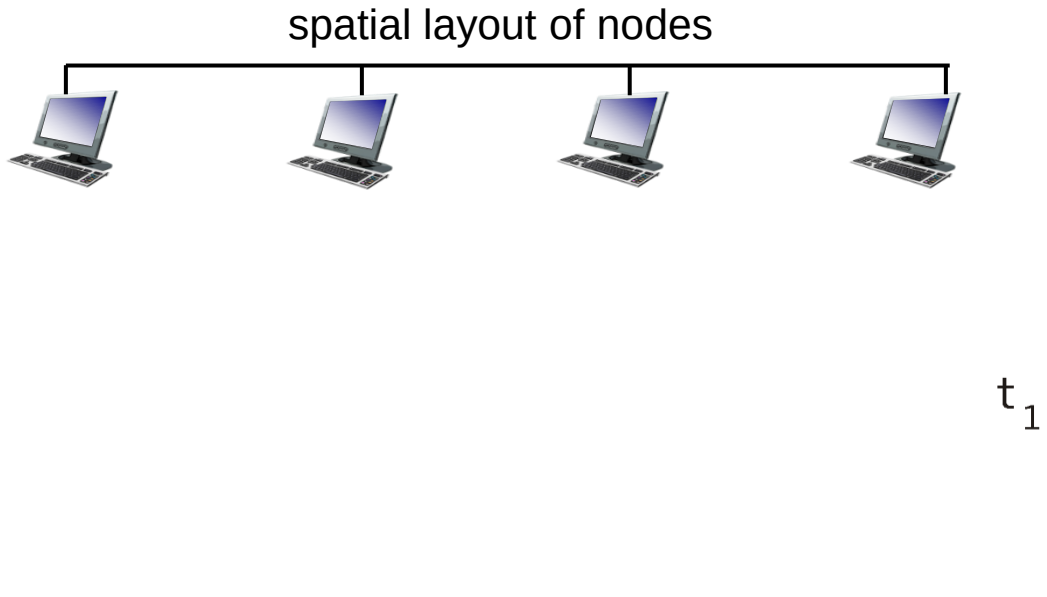
CSMA发送冲突的例子

- 一个广播总线连接4个节点(A、B、C、D)传输的时空图。



- 时间t_0：节点B侦听到信道空，开始传输帧，沿着媒体传播比特。
- 时间t_1（$t_1 > t_0$）：节点D有帧要发送。B的传输信号未到D，D检测到信道空，开始传输。很快，B的传输开始在D节点干扰D的传输（冲突）

端到端信道传播时延：信号从一个节点到另一个节点所花费的传播时间。传播时延越长，节点不能侦听到另一个节点已经开始传输的可能性越大。

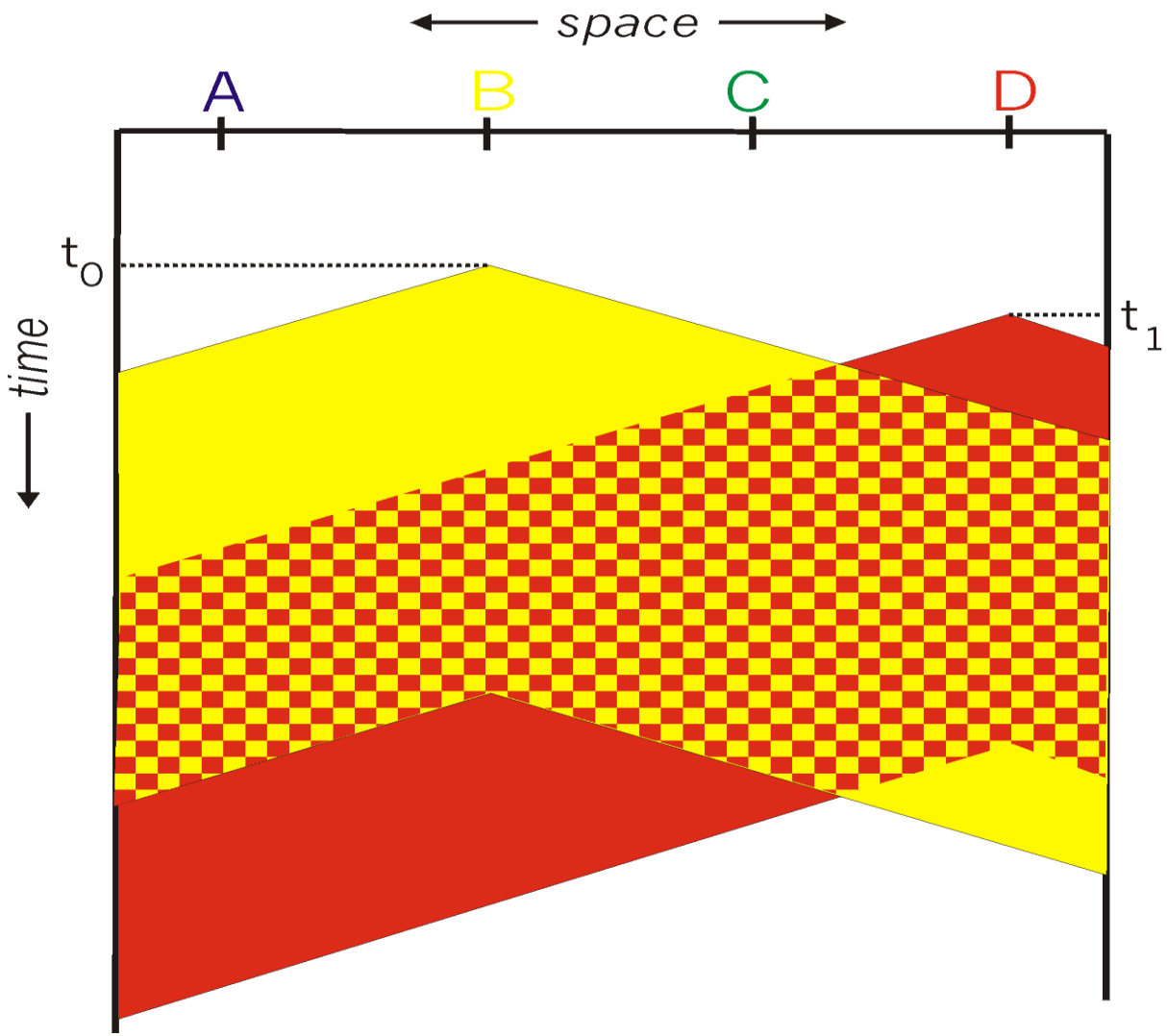


带来问题：信道浪费

- 节点没有进行冲突检测，既使发生了冲突，节点仍继续传输它们的帧。但该帧已经被破坏、是无用的帧，信道传输时间被浪费。

注意：

距离与传播时延对碰撞概率的影响。



带冲突检测的CSMA(CSMA/CD)

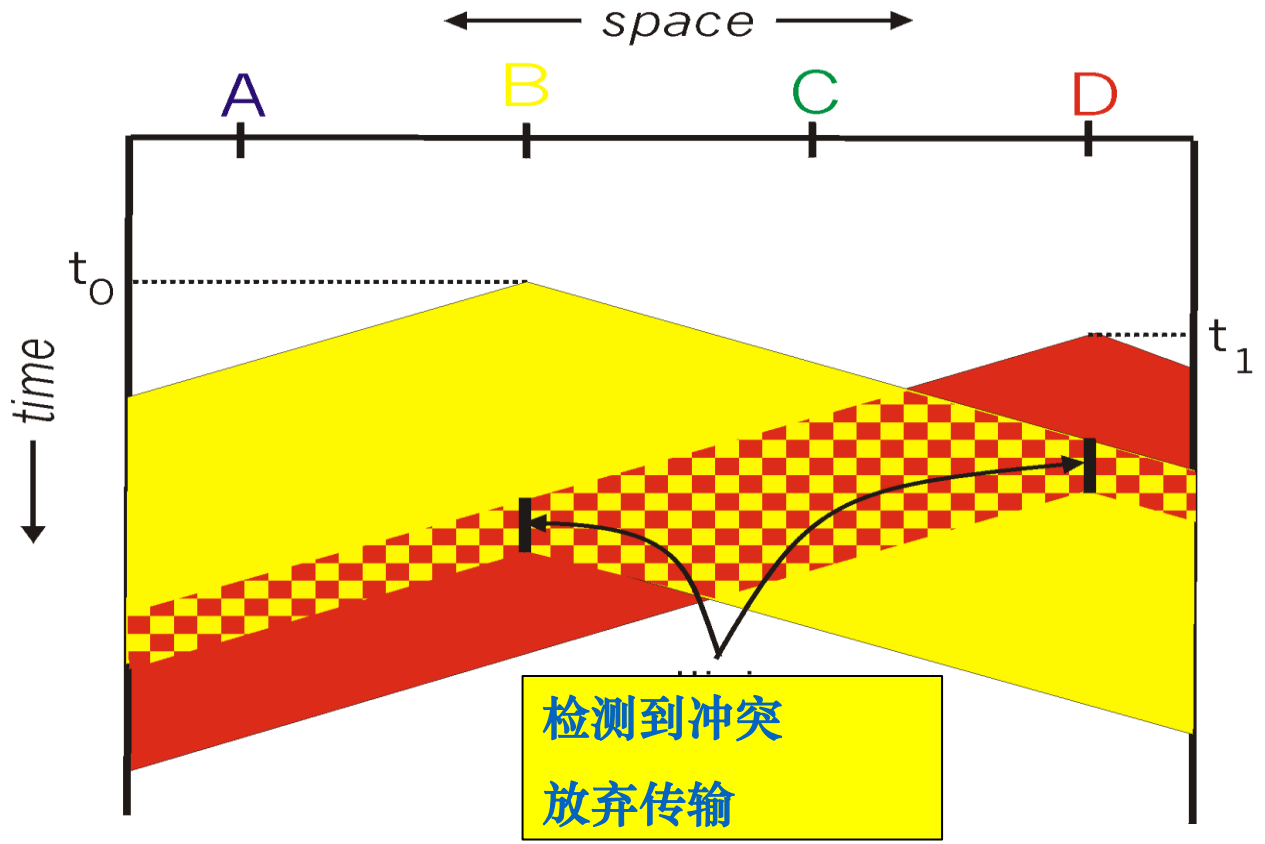
增加“载波侦听”和“冲突检测”两个规则。

- 基本原理： 传送前侦听
  - 信道忙：延迟传送
  - 信道闲：传送整个帧
- 发送同时进行冲突检测：一旦检测到冲突就立即停止传输， 尽快重发。
- 目的：缩短无效传送时间，提高信道的利用率。

CSMA/CD的例子

两个节点B、 D在检测到冲突之后很短的时间内都放弃传输。

以太网即采用CSMA/CD协议。



以太网CSMA/CD的运行机制

1. 适配器从网络层得到分组, 创建帧

2. 如果适配器侦听到信道空闲,开始传送帧。如果信道忙, 它会等到信道空闲才传送帧

3. 如果适配器传送整个帧时，都没有检测到其它传输, 则完成该帧的传送

4. 如果适配器在发送中检测到其它传送,就放弃传送，并发送一个拥塞信号

5. 放弃传送后，适配器进入指数回退阶段，即该帧经过n次冲突后，适配器在{0,1,2,…,2^m-1}中随机选取一个K值，其中m=min(n,10),然后等待K*512比特时间后,回到第2步

二进制指数回退算法

以太网CSMA/CD的运行机制讨论

拥塞信号

- 48比特，确保所有传送者知道冲突发生

比特时间

- 对于10 Mbps Ethernet 为0.1微秒，当K=1023，等待时间大约50毫秒

二进制指数回退算法

- 目标：适配器依据当前负载情况重传，重负载时等待时间变长
- 第一次冲突： 在{0,1}中选k值；延迟Kx512比特时间传送
- 第二次冲突：在{0,1,2,3}中选k值…
- 10次以后，在 {0,1,2,3,4,…,1023}中选k值。
- K是等概率选择

6.3.3 轮流协议

- **多路访问协议理想特性：**
  - 只有一个节点活动时，吞吐量R b/ s；
  - 有M个节点活动时，吞吐量R/M b/ s。
  - ALOHA和CSMA协议有第一个特性，但没有第二个特性

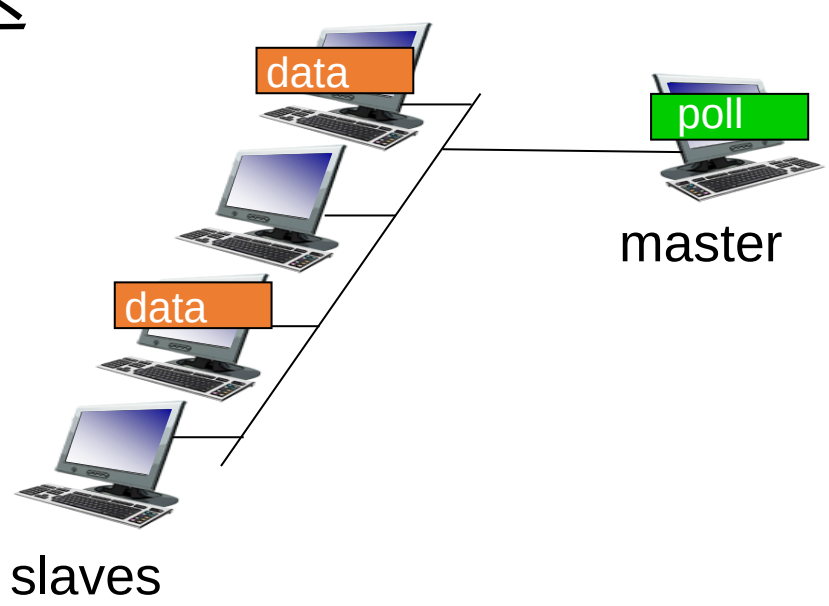
1、轮询协议

2、令牌传递协议

轮流协议

轮询(polling):

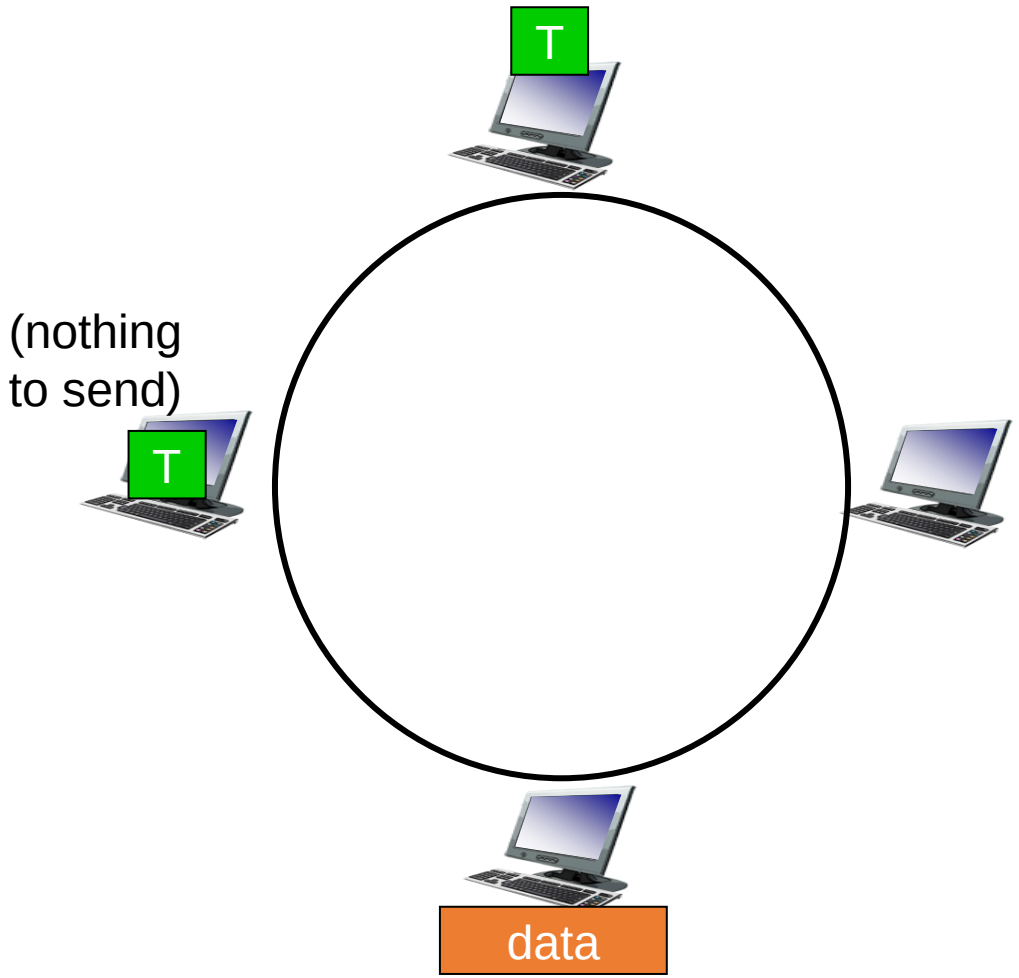
- 主节点“邀请”从节点依次传送
- 问题:
  - 轮询的开销
  - 延时
  - 单点故障(主节点)



轮流协议

令牌传递(token passing):

- 控制令牌顺序从一个节点传递到下一个节点。
- 问题:
  - 令牌开销
  - 延时
  - 单点失效(token)



多路访问控制协议的总结

- **信道划分**：时分，频分，码分
- **随机接入**：
  - ALOHA, S-ALOHA, CSMA, CSMA/CD
  - 载波侦听：在某些技术中容易实现(有线)，在有些技术中比较困难(无线)
  - CSMA/CD used in Ethernet
  - CSMA/CA used in 802.11
- **轮流**
  - 来自中心站的轮询
  - 令牌传递

5.4 交换局域网

局域网概述

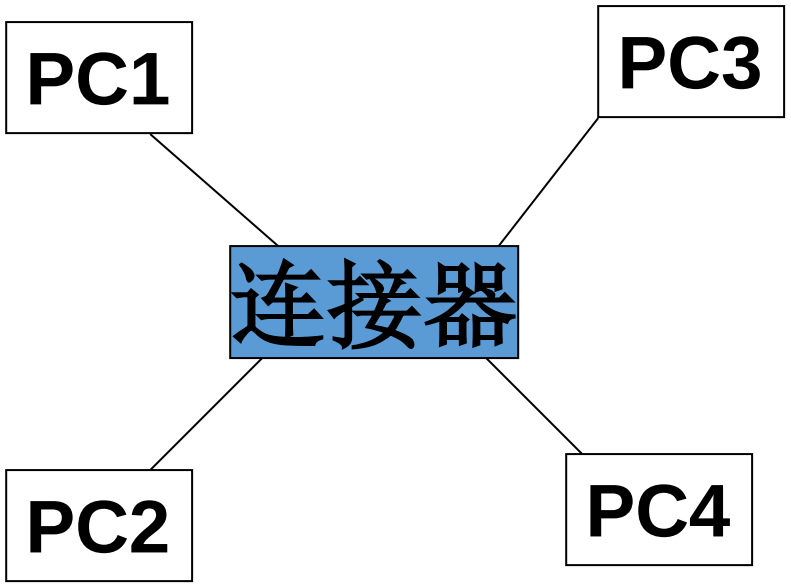
- 局域网：Local Area Network（LAN）
- 多址访问协议广泛应用于局域网
- 基于随机访问的CSMA/CD广泛应用于局域网
- 基于令牌传递技术的令牌环和FDDI在局域网技术中变得次要或被淘汰
- 链路层技术的发展，使得局域网、城域网、广域网的概念变得越来越模糊和不重要

局域网概述

- 主要特点：网络为一个组织所拥有，且地理范围和站点数目均有限
- 局域网按拓扑结构进行分类：星形网、环形网、总线网、树形网和网状网

常见的网络拓扑结构

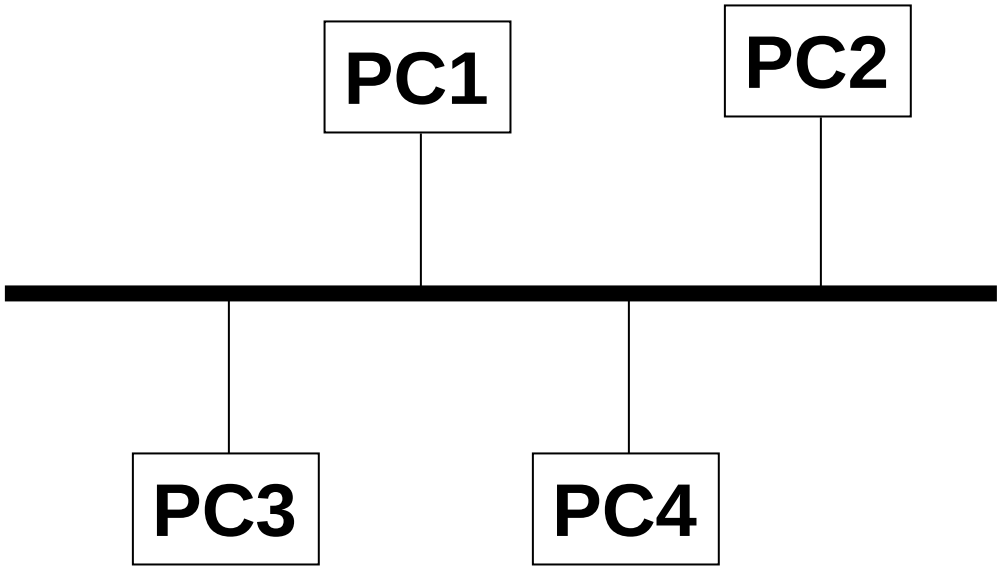
- 星型结构



- 辐射状连接
- 中央结点集中式通信控制
- 优点：结构简单，访问协议简单，单个节点的故障不会影响到整个网络。
- 缺点：对中央结点的可靠性要求很高，一有故障，全网瘫痪。

常见的网络拓扑结构

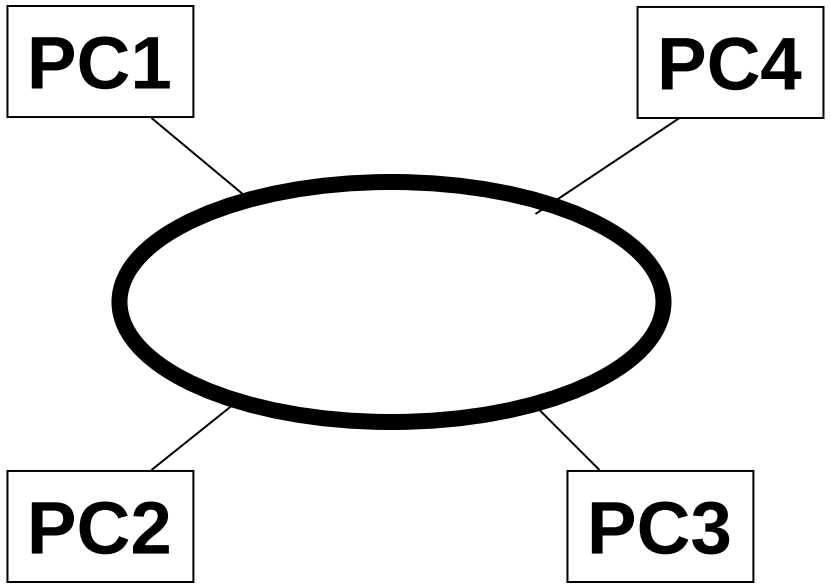
- 总线结构



- 所有的站点都连接在同一根传输线，即“总线”上
- 优点：结构简单，易于扩充
- 缺点：故障检测比较困难

常见的网络拓扑结构

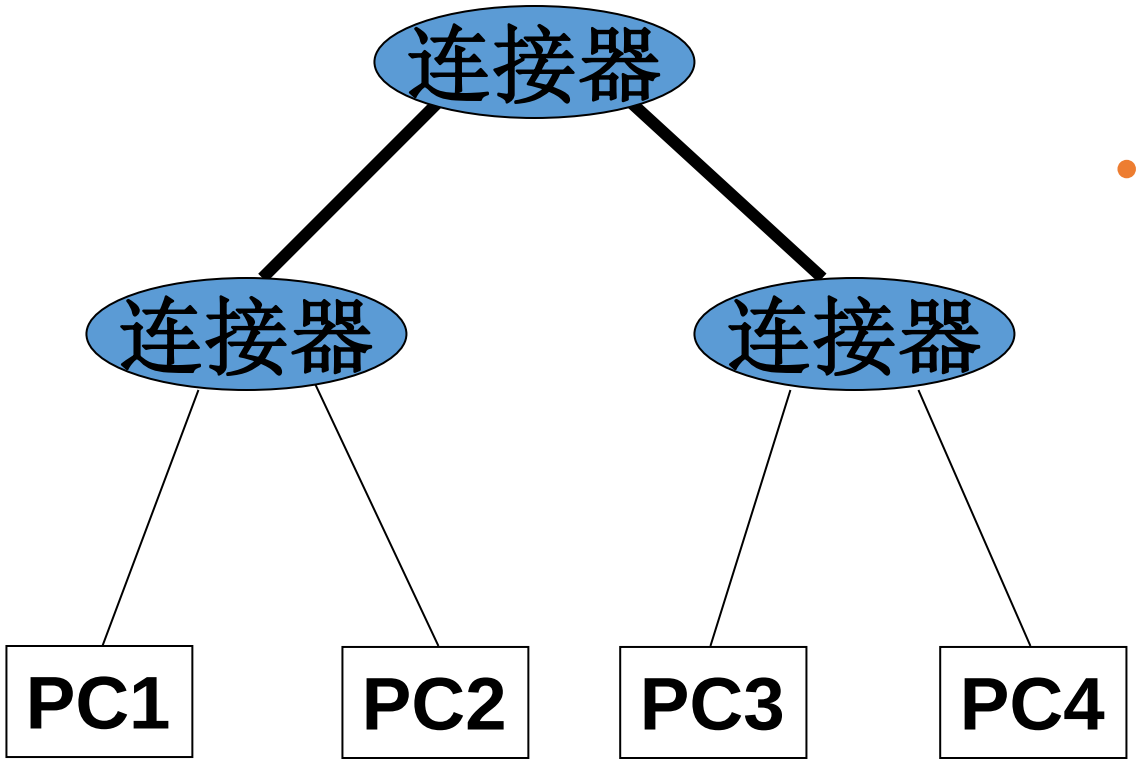
- 环型结构



- 站与站点之间首尾相接，形成一个环，数据只能沿单方向传输
- 优点：这种结构适合于光纤介质。实时性较强
- 缺点：如果处理不当，站点的故障会引起全网故障

常见的网络拓扑结构

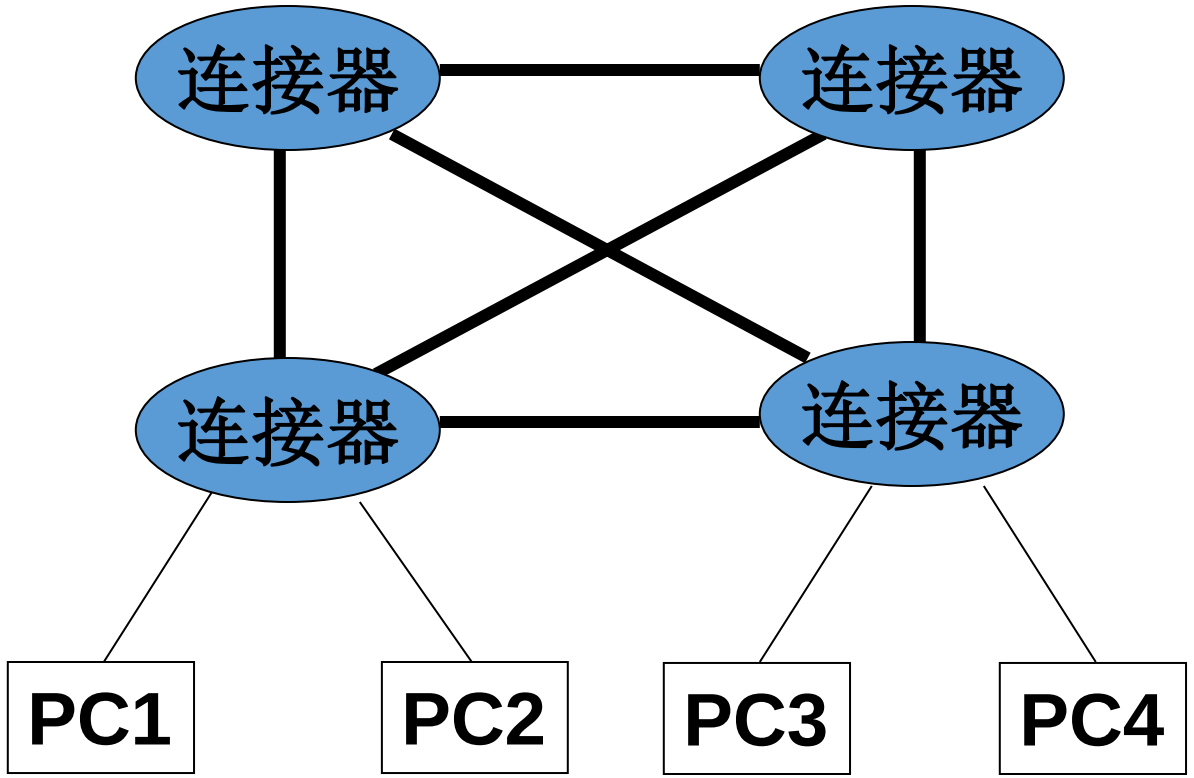
- 树型结构



- 它是从星型拓扑演变而来的，形状像一棵倒挂的树
- 特点：与星型拓扑大致相似。它与星型结构相比降低了通信线路成本，增加了网络复杂性

常见的网络拓扑结构

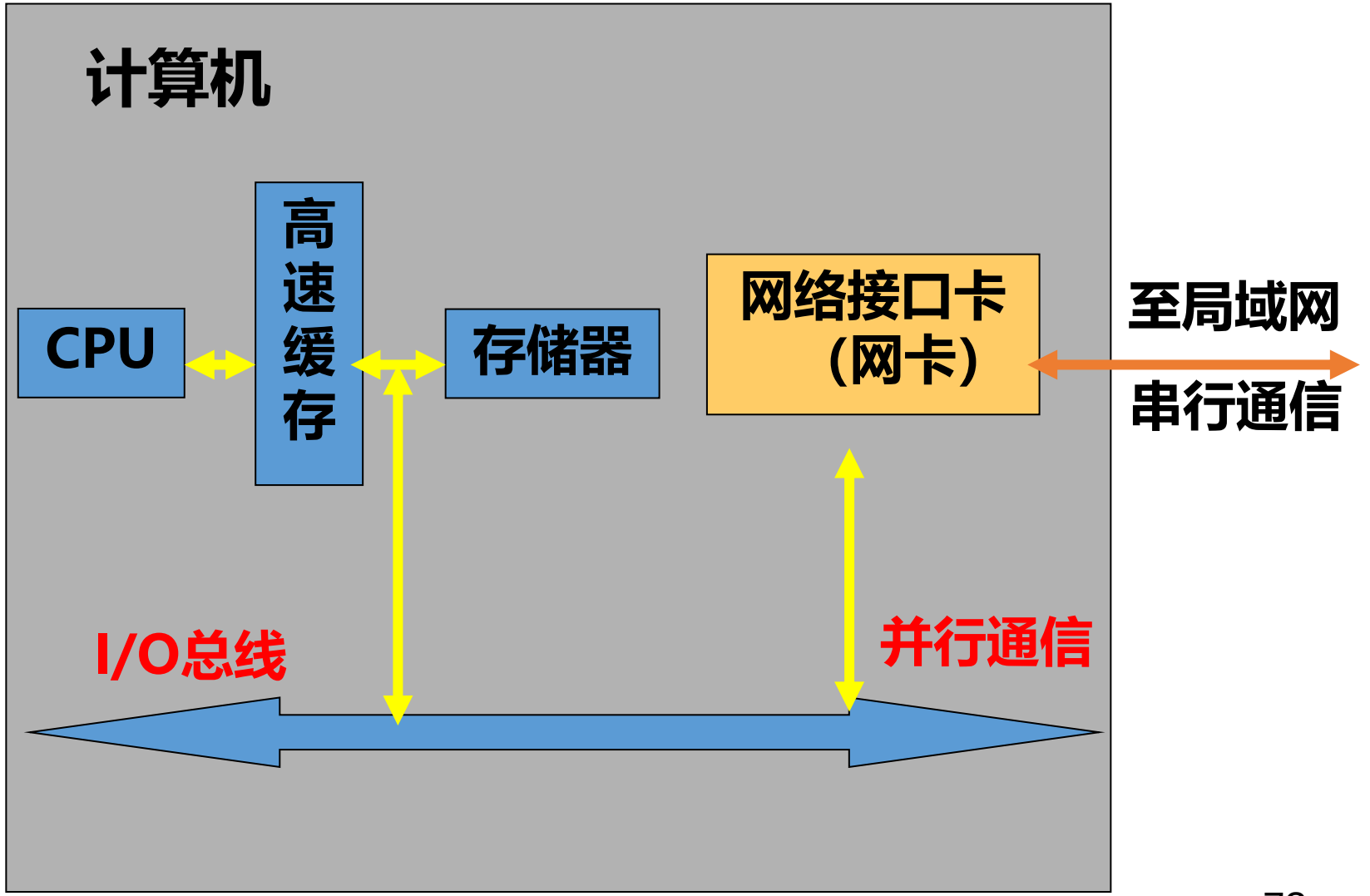
- 网状结构



- 网状网络的每一个站点都与其它站点一一直接互连
- 优点：连接方法主要是利用冗余的连接，实现站与站之间的高速传输和高容错性能，以提高网络的速度和可靠性
- 优点：关系复杂，建网难，维护难

计算机与局域网的连接

- 计算机与局域网通过网络接口板进行连接，网络接口板又称通信适配器（Adapter）或网络接口卡NIC（Network Interface Card），通常我们称为“网卡”。

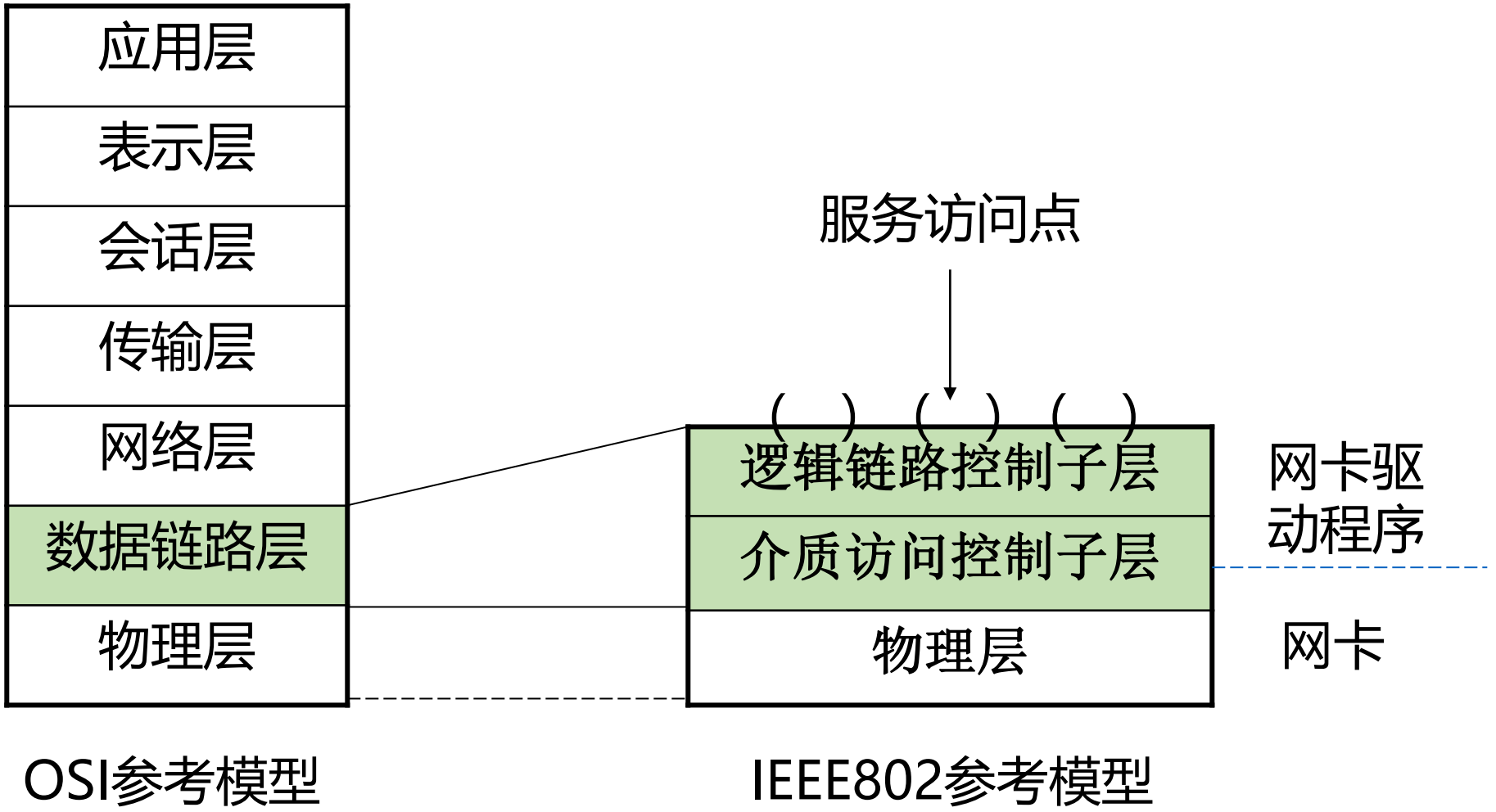


局域网体系结构

- 美国电气与电子工程师协会IEEE 802委员会制定的IEEE 802局域网标准已经得到了国际标准化组织ISO的采纳，成为计算机局域网的事实标准
- IEEE 802局域网参考模型是针对局域网的网络体系结构特点而制定的，它遵循ISO/OSI参考模型的原则，解决物理层和数据链路层的功能以及与网络层的接口服务、网际互连的高层功能

局域网体系结构

IEEE802与OSI参考模型对应关系



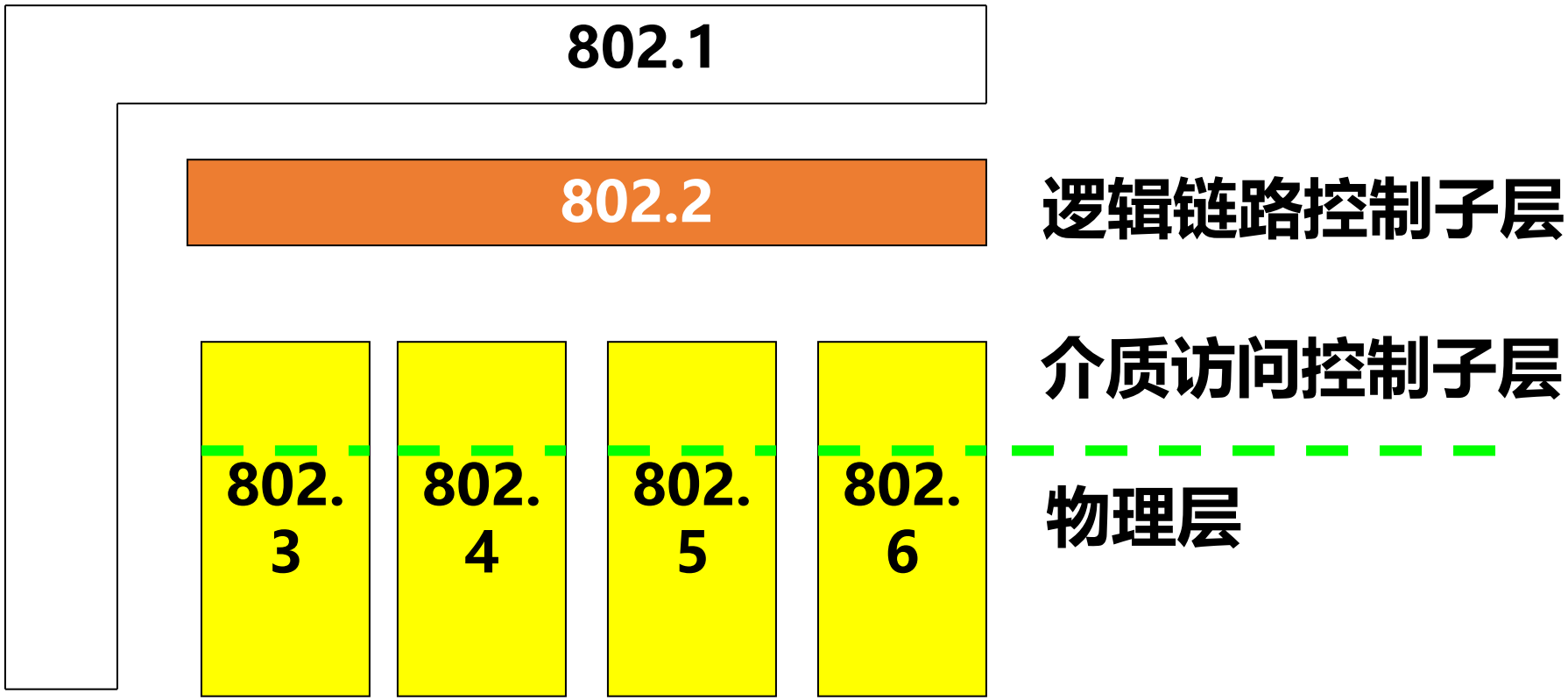
局域网体系结构

IEEE802网络协议标准(部分)

- IEEE802.1　体系结构及高层接口
- IEEE802.2　逻辑链路控制LLC
- IEEE802.3　CSMA/CD（带冲突检测的载波侦听多路复用）访问控制方法和物理层技术规范
- IEEE802.4　令牌总线访问控制方法和物理技术规范
- IEEE802.5　令牌环访问控制方法和物理层规范
- IEEE802.6　城域网访问控制方法和物理层技术规范
- IEEE 802.10：网络安全技术咨询组，定义了网络互操作的认证和加密方法。
- IEEE 802.11：无线局域网（WLAN）的介质访问控制协议及物理层技术规范 (a,b,…,ae)
- IEEE 802.15：无线个人网（Wireless Personal Area Networks，WPAN）技术规范，比如蓝牙

局域网体系结构

几个基础性IEEE802标准的关系



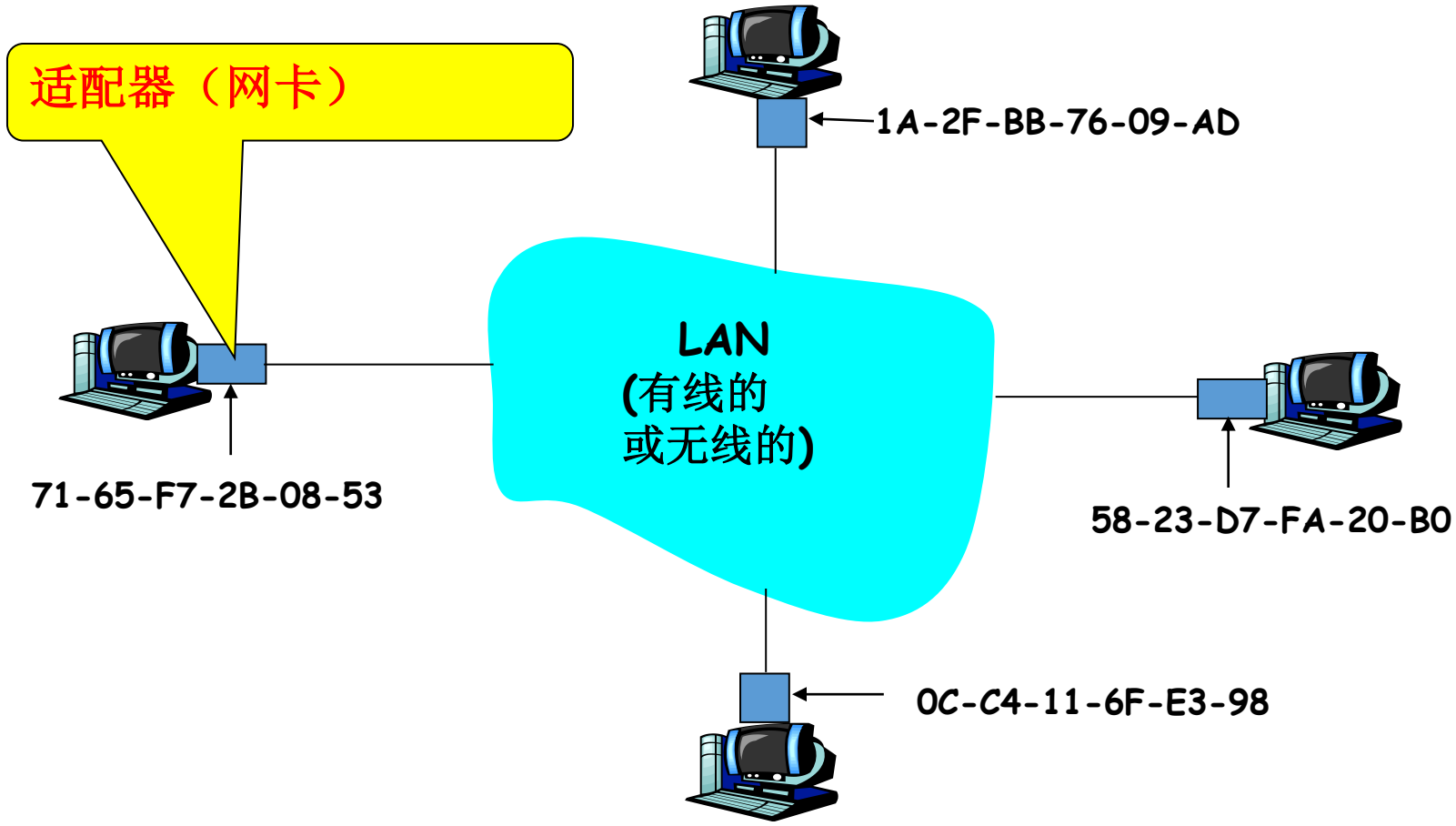
5.4.1 链路层寻址和ARP

每个节点有网络层地址和链路层地址。

- 网络层地址：节点在网络中分配的一个唯一地址（IP地址）。用于把分组送到目的IP网络。长度为32比特（IPv4）。
- 链路层地址：又叫做MAC地址或物理地址、局域网地址。
  - 用于把数据帧从一个节点传送到另一个节点(同一网络中)。
- MAC地址（LAN地址、物理地址）：
  - 节点“网卡”本身所带的地址（唯一）。
  - MAC地址长度通常为6字节(48比特)，共2^{48}个。 1A-2F-BB-76-09-AD
  - 6字节地址用16进制表示，每个字节表示为一对16进制数
  - 网卡的MAC地址是永久的（生产时固化在其ROM里）

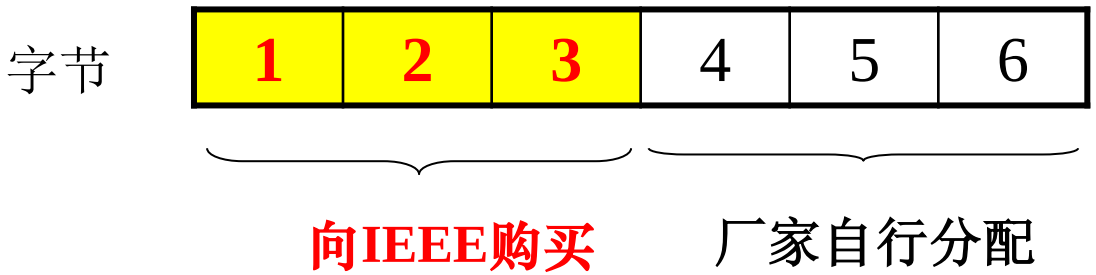
局域网地址

局域网中每个网卡都有唯一的局域网地址



MAC地址分配

- 由专门机构IEEE管理物理地址空间
  - 负责分配六个字节中的前三个字节（高24位，地址块）
- MAC 地址是平面结构
  - 带有同一网卡的节点，在任何网络中都有同样的MAC地址。
- IP地址具有层次结构
  - 当节点移动到不同网络时，节点的IP地址发生改变。

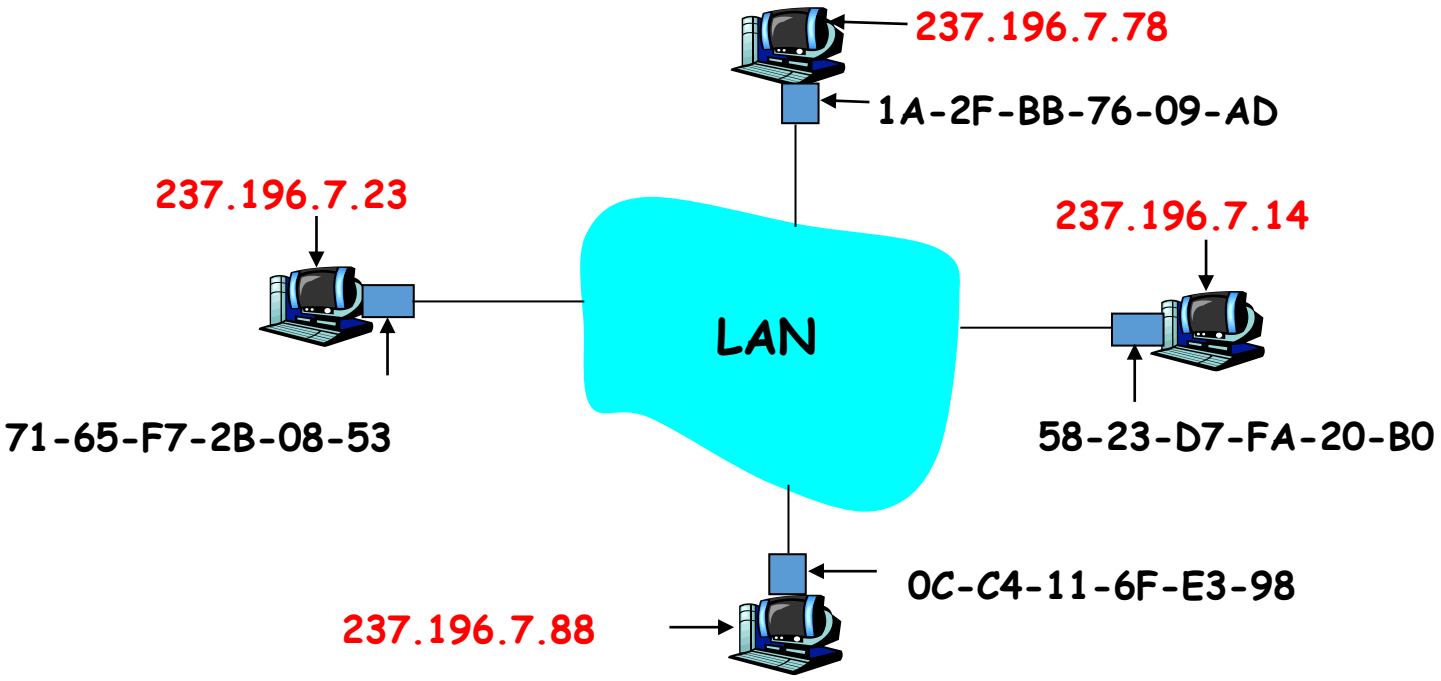


MAC地址识别

- 广播信道的局域网中，一个节点发送的帧，在信道上广播传输，其他节点都可能收到该帧。
- 大多数情况，一个节点只向某个特定的节点发送。
- 由“网卡”负责MAC地址的封装和识别。
- **发送适配器**：将目的MAC地址封装到帧中，并发送。所有其他适配器都会收到这个帧。
- **接收适配器**：检查帧的目的MAC地址是否与自己MAC地址相匹配：
  - 匹配：接收该帧，取出数据报，并传递给上层。
  - 不匹配：丢弃该帧。
- 广播帧：发送给所有节点的帧　　全1地址：FF-FF-FF-FF-FF-FF

回顾1：节点的3种不同地址表示

- 应用层的主机名、网络层的IP地址和链路层的MAC地址
- 实际在链路上传输时，根据MAC地址，确定相应的节点



回顾2：地址之间的转换

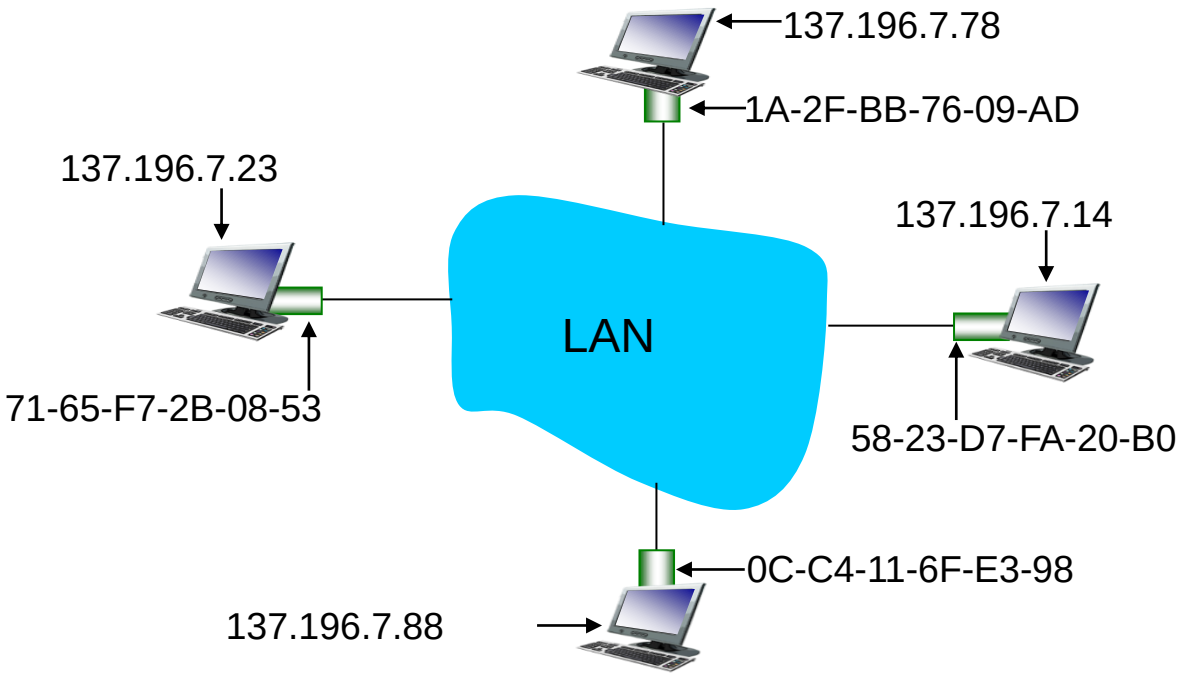
通信时，需要进行地址转换：

主机名 → IP地址 → MAC地址

- DNS域名系统：将主机名解析到IP地址。
  - DNS为在因特网中任何地方的主机解析主机名。
- ARP地址解析协议：将IP地址解析到MAC地址。
  - ARP只为在同一个LAN上的节点解析IP地址。

ARP: 地址解析协议

问题：如何根据一个主机的IP地址，查找其MAC地址



ARP表： 局域网上的每个节点（主机、路由器）都有这个表

- 为某些局域网节点进行IP/MAC地址映射：

  < IP address; MAC address; TTL>

- TTL (存活时间): 地址映射将被删除的时间（通常为20分钟）

ARP: 两个主机位于同一个局域网

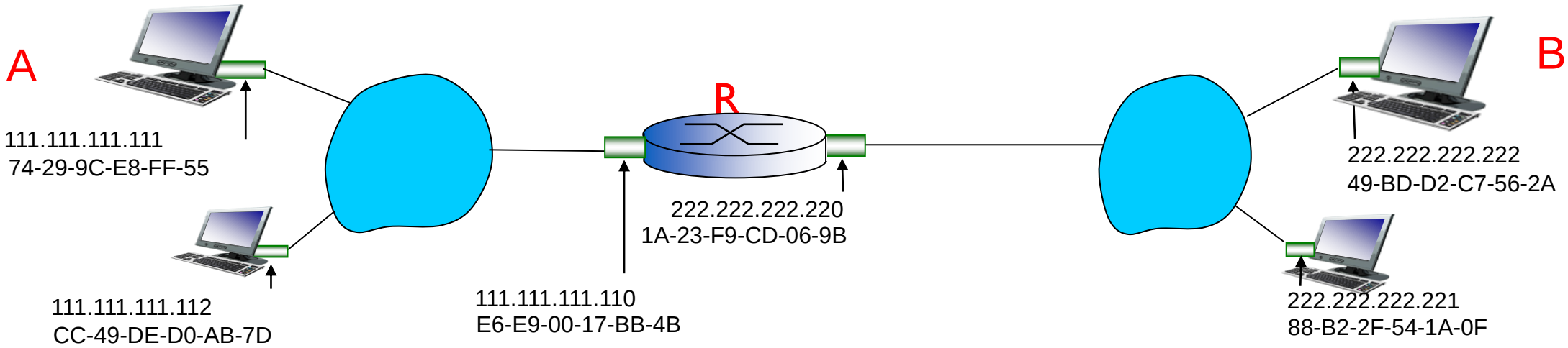
- 主机A希望发送数据报给主机B
  - B的MAC地址不在A的ARP映射表中
- 主机A 广播 ARP查询分组, 其中包含B的IP地址
  - 目的MAC地址 = FF-FF-FF-FF-FF-FF
  - 局域网中所有节点收到ARP查询分组
- 主机B收到ARP查询分组，返回B的MAC地址给主机A
  - 包含有B的MAC地址的帧发送给主机A(单播)

- 主机A在它的ARP表中缓存 **IP-to-MAC 地址对，**直到信息
  - 软状态：信息超时会被删除，除非有新的更新消息
- ARP是即插即用的:
  - 节点创建ARP表不需要网络管理员的干预

发送数据报到子网以外

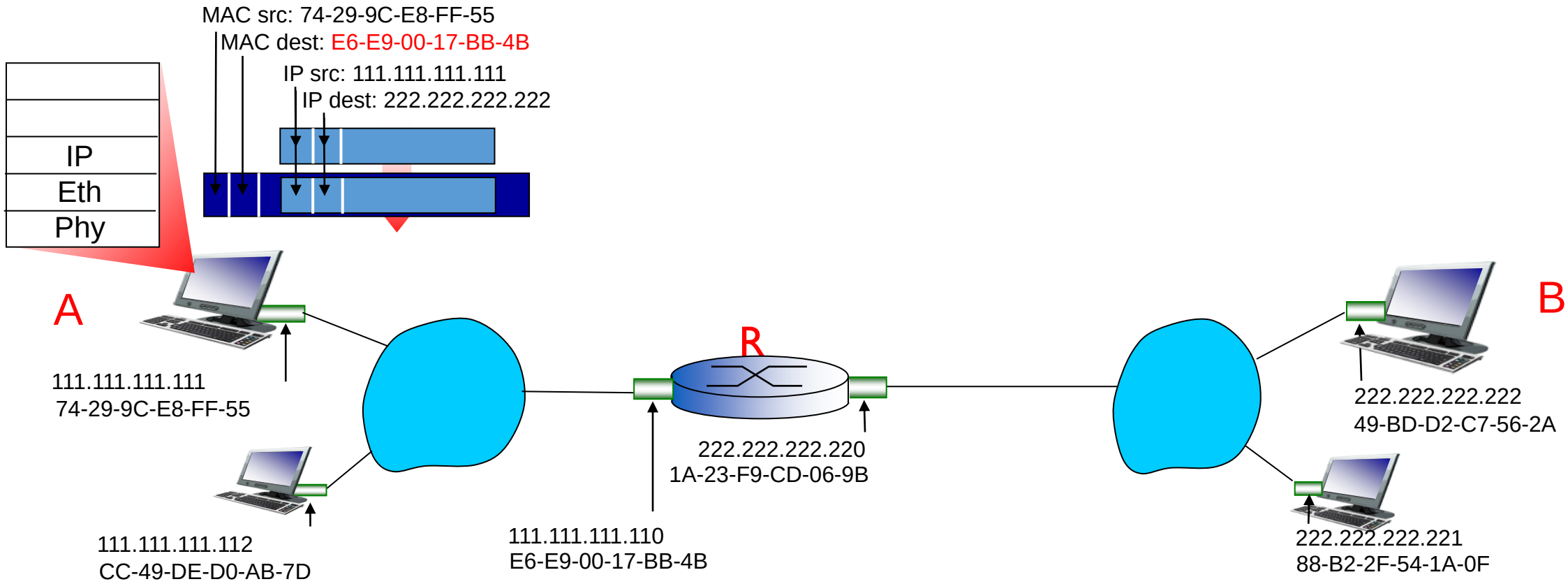
示例: 主机A经路由器R发送数据报给主机B

- 集中在寻址上——IP层(数据报)和MAC层(数据帧)
- 假设主机A知道主机B的IP地址
- 假设主机A知道第一跳路由器R的IP地址(通过DHCP协议)
- 假设主机A知道路由器R的MAC地址(通过ARP协议)



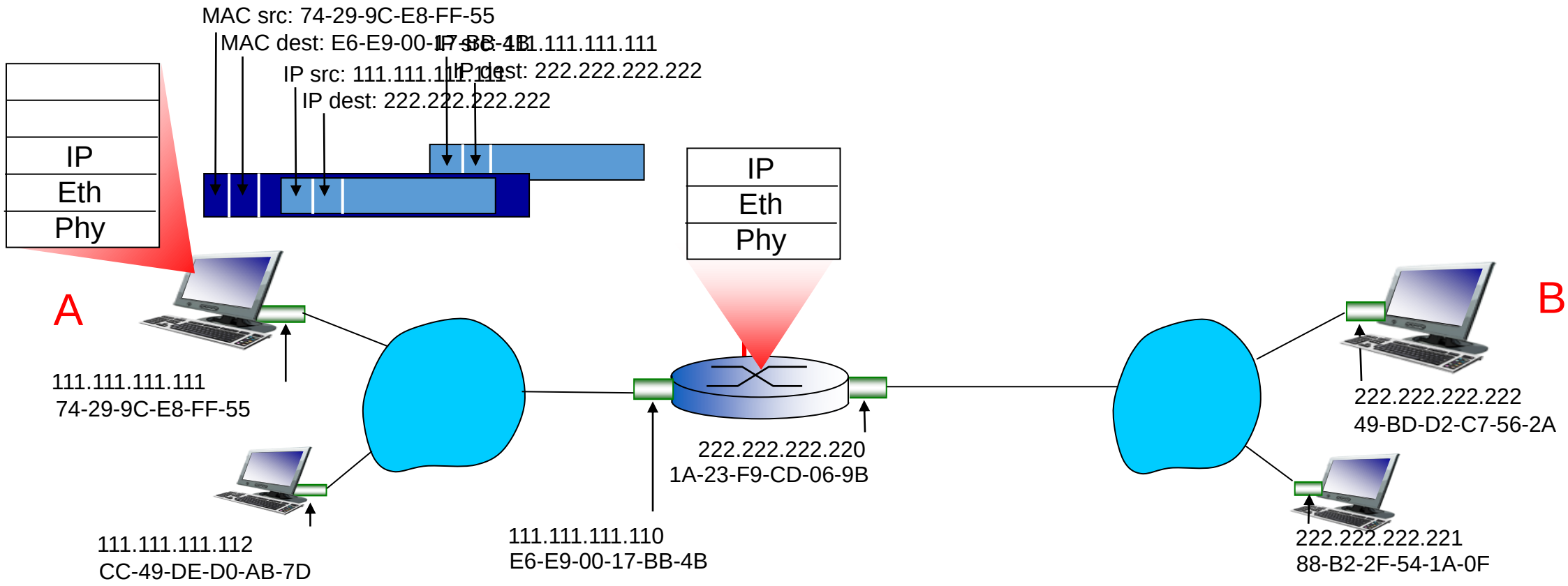
发送数据报到子网以外

- 主机A构建IP数据报，源地址是A的IP地址，目的地址是B的IP地址
- 主机A构建链路层数据帧，目的MAC地址是路由器左边端口的MAC地址



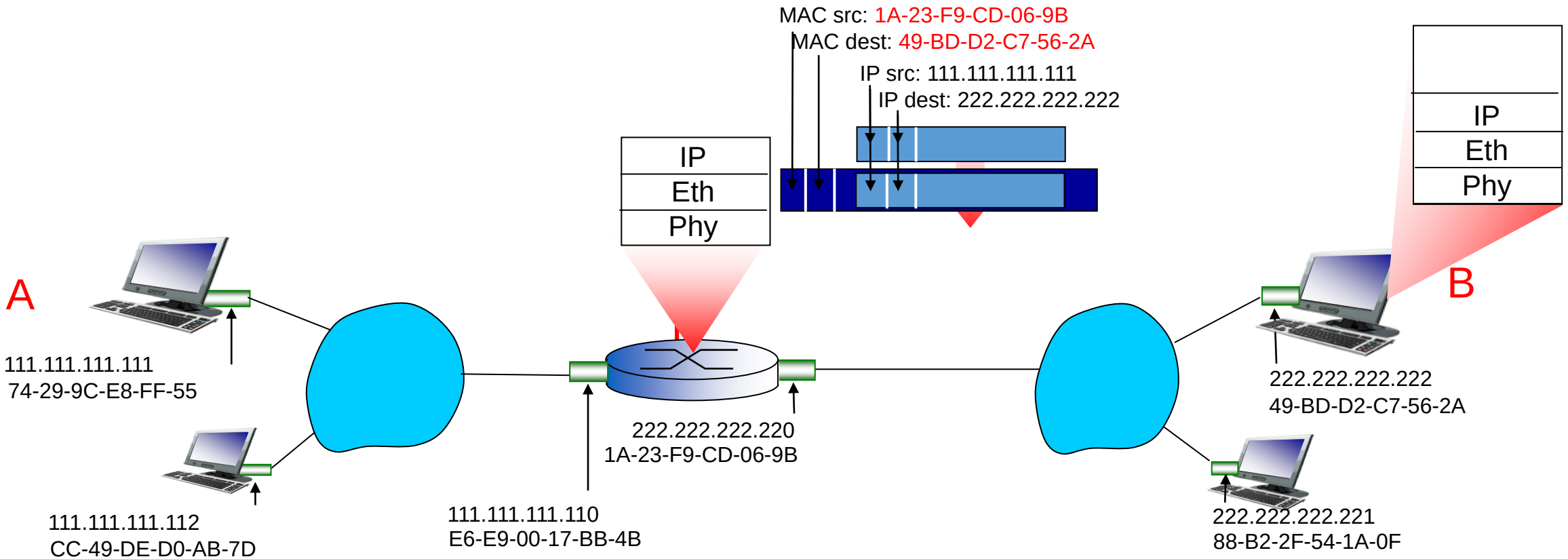
发送数据报到子网以外

- 数据帧从主机A发送到路由器R
- 路由器R收到数据帧，抽取出数据报递交到IP层



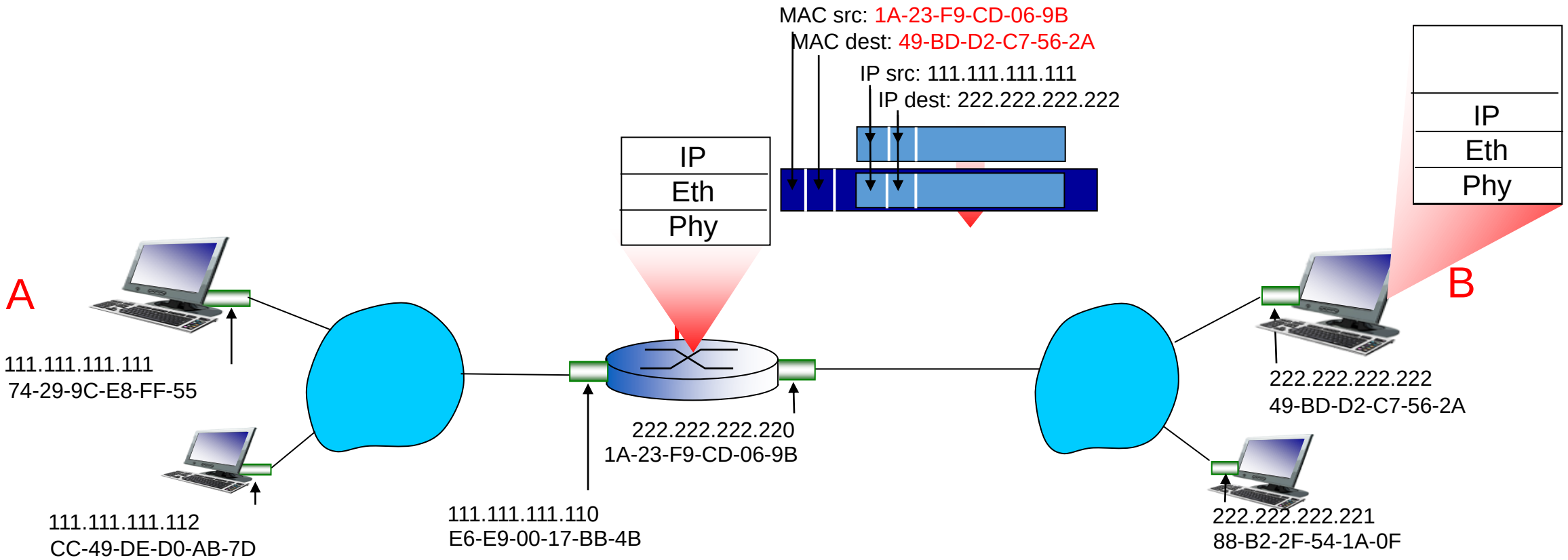
发送数据报到子网以外

- 路由器R转发数据报，源地址为A的IP地址，目的地址为B的IP地址
- 路由器R将该数据报封装成链路层帧，目的MAC地址为主机B的MAC地址



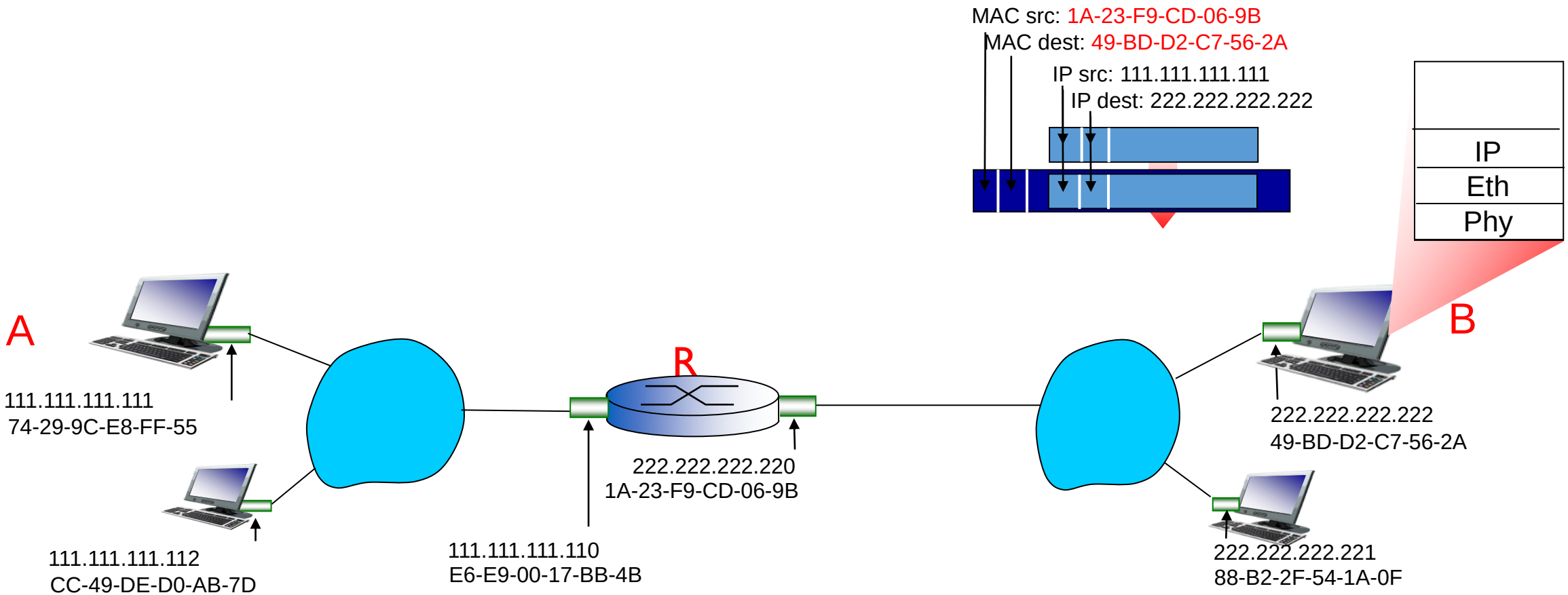
发送数据报到子网以外

- 路由器R转发数据报，源地址为A的IP地址，目的地址为B的IP地址
- 路由器R将该数据报封装成链路层帧，目的MAC地址为主机B的MAC地址



发送数据报到子网以外

- 路由器R转发数据报，源地址为A的IP地址，目的地址为B的IP地址
- 路由器R将该数据报封装成链路层帧，目的MAC地址为主机B的MAC地址

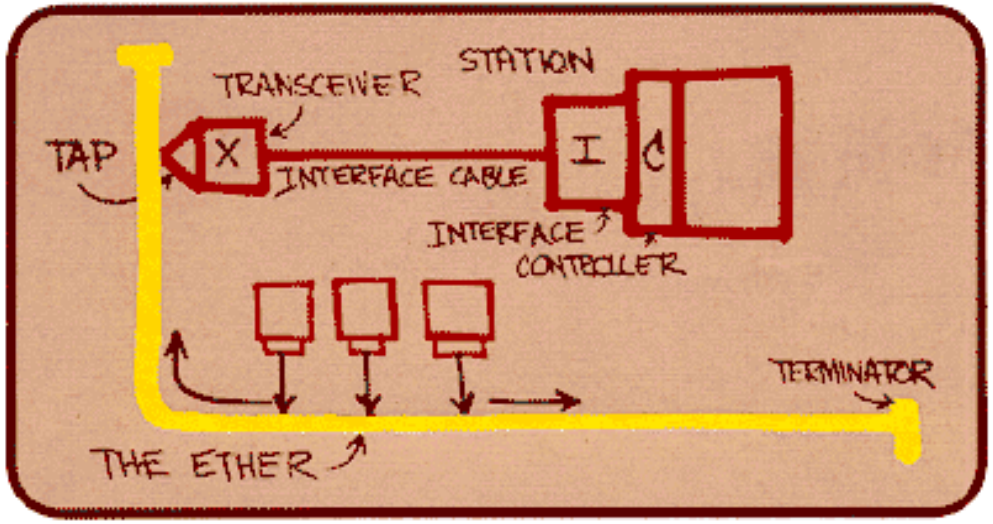


5.4.2 以太网(Ethernet)

到目前为止，以太网是最为著名的有线局域网技术

以太网成功的原因：

- 是第一个广泛使用的局域网技术；
- 简单、便宜；
- 版本不断更新，数据速率更高、成本更低。



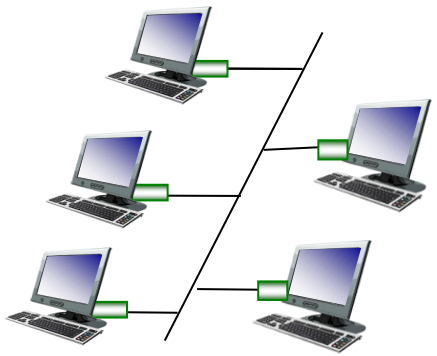
Metcalfe’s Ethernet sketch

以太网

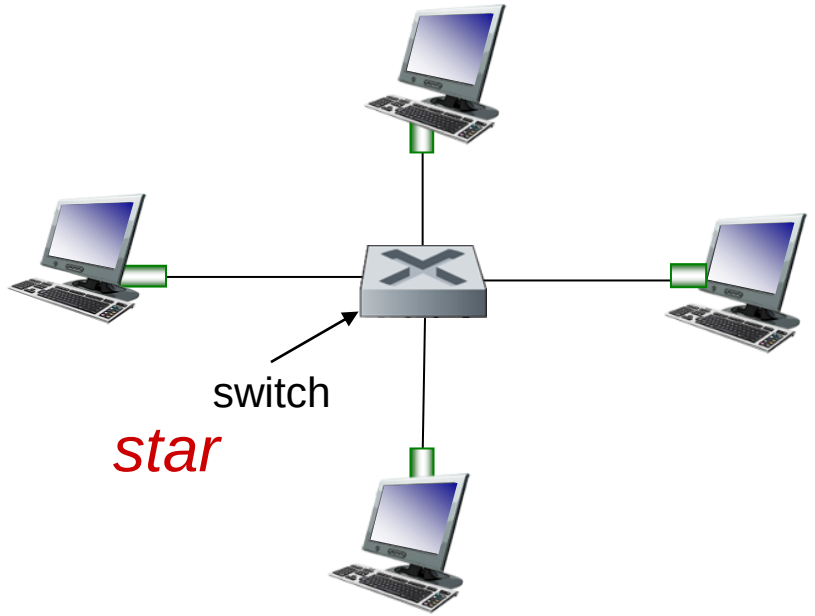
- 以太网是70年代由Digital Equipment 、Intel和Xerox三家公司开发的局域网组网规范，并于80年代初首次出版，称为DIX1.0。1982年修改后的版本为DIX2.0。不久这三家公司公布了与IEEE802.3一致的以太网规范。Ethernet和IEEE802.3虽然有很多规定不同，但术语Ethernet通常认为与802.3是兼容的。IEEE将802.3标准提交国际标准化组织(ISO)第一联合技术委员会(JTC1),再次经过修订变成了国际标准ISO802.3。
- 最初的以太网是采用同轴电缆来连接各个设备的，如今则广泛使用双绞线、光纤等
- 如今的无线局域网wifi也使用了若干以太网的技术和规范

以太网的物理拓扑结构

- 总线(bus): 一直流行到90年代中期
  - 所有节点都属于相同的冲突域
- 星形(star):目前流行
  - 中心是交换机
  - 每个端口运行一个独立的以太网协议(节点相互之间不发生碰撞)



bus: coaxial cable



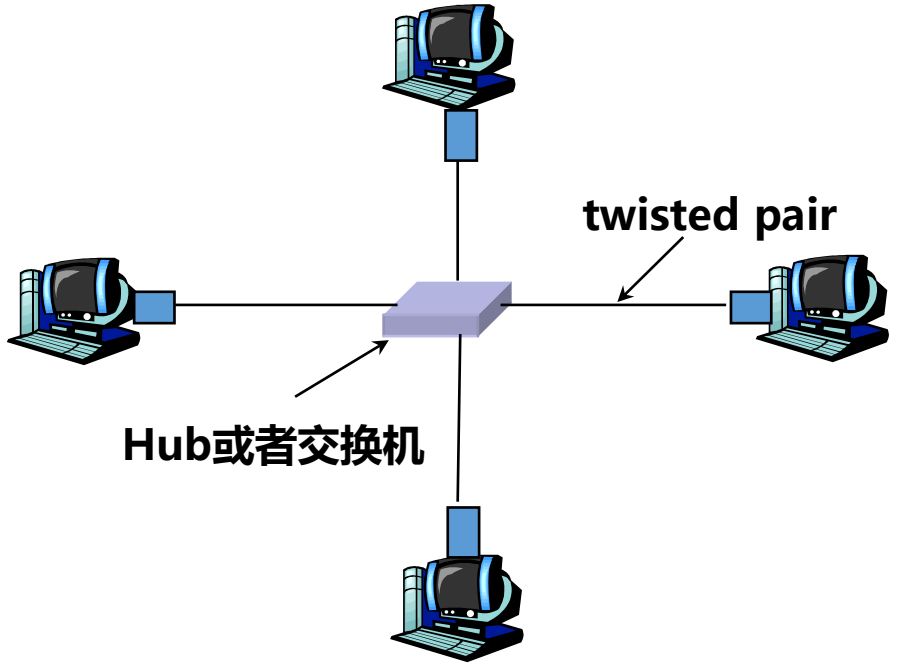
star

以太网物理层标准

- 以太网的物理层标准（与传输介质相对应）：
  - 10BASE-T
  - 10BASE2
  - 10BASE5
  - 100BASE-T
  - 1000BASE-T
  - 1000BASE-LX
  - 1000BASE-SX等

以太网技术:10Base-T 和 100-BaseT

- 10/100 Mbps速率；后者被称为 “快速ethernet”
- T 表示双绞线
- 各节点都连接到集线器上 “星型拓扑结构”；在节点和适配器间最大距离为100米

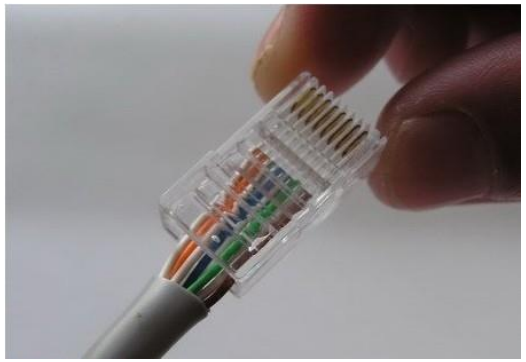


以太网技术:10-BaseT 和 100-BaseT

- 双绞线（Twisted Pair）是由两条相互绝缘的导线按照一定的规格互相缠绕（一般以逆时针缠绕）在一起而制成的一种通用配线，属于信息通信网络传输介质。双绞线过去主要是用来传输模拟信号的，但现在同样适用于数字信号的传输。
- 分类：
  - 1类线(CAT1)~7类线(CAT7)，目前常用的是5类和超五类
  - 按是否有屏蔽层，分为屏蔽线和非屏蔽线

以太网技术:10Base-T 和 100Base-T

- 双绞线的两头需要按一定顺序把各根线排好，然后接入水晶头中方可使用
- 国际上常用的制作双绞线的标准包括EIA/TIA 568A和EIA/TIA 568B两种



以太网技术:1000Base系列

- 1000BASE系列有四种传输介质标准:
  - 1000BASE-T

    使用非屏蔽双绞线作为传输介质提供1000M的传输速度
  - 1000BASE-LX

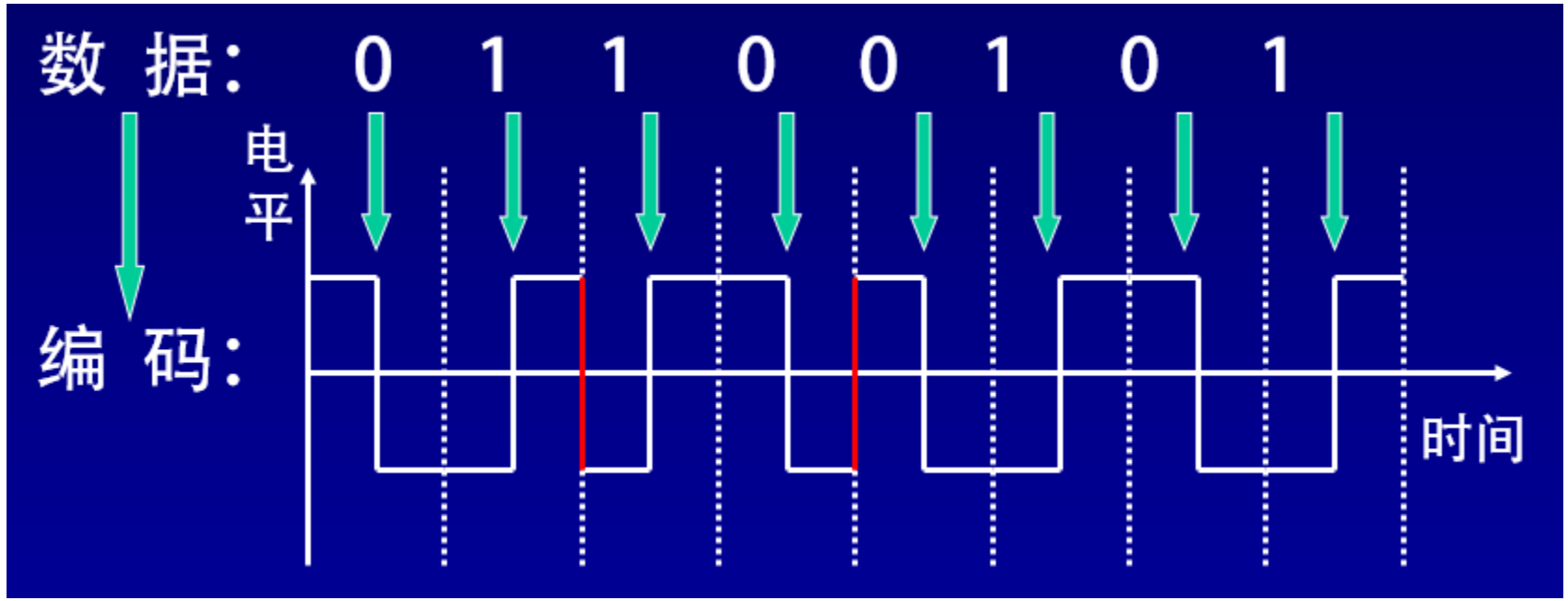
    使用单模光纤
  - 1000BASE-SX

    使用多模光纤
  - 1000BASE-CX

    使用平衡、屏蔽铜缆，它可以用于机房的互连

以太网物理层标准

- 曼彻斯特编码（10BASE - T物理层编码）*



以太网物理层标准

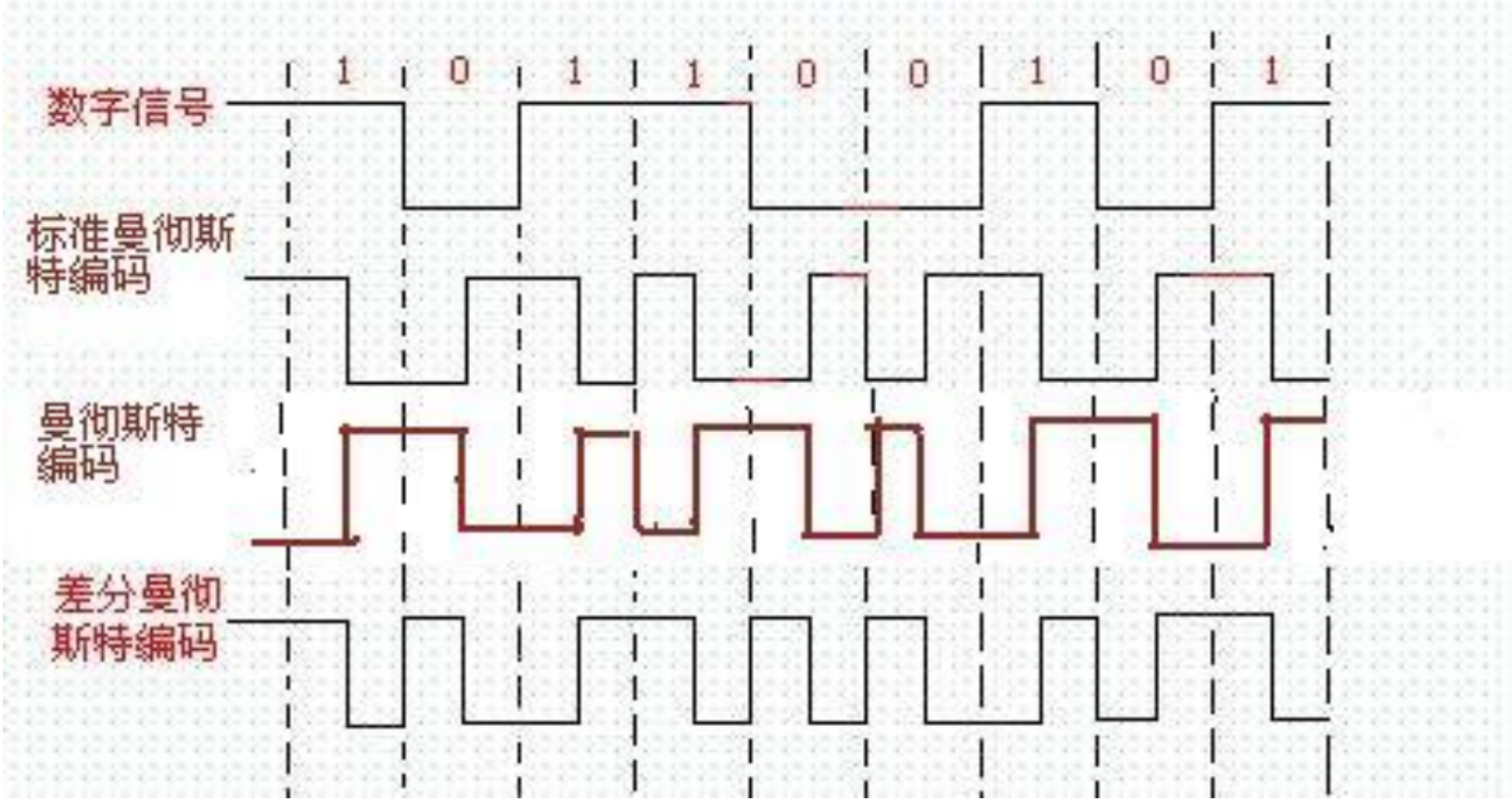
- 曼彻斯特编码机制 *
  - 利用一个时钟周期中间位置的电平跳变来表示“0”和“1”。
  - 中间跳变是由低电平跳变到高电平表示“1”
  - 中间跳变是由高电平跳变到低电平表示“0”
  - 注：在时钟周期边界处可以任意跳变而不影响编码含义。
  - 另有一种称为“标准曼彻斯特编码”的机制，跳变方法和上面刚好相反

以太网物理层标准

- 差分曼彻斯特编码机制 *
  - 在信号位开始时不改变信号极性，表示1;
  - 在信号位开始时改变信号极性，表示0;
  - 每个信号位中间都要跳变;
  - 在第一个信号时:
    - 如果中间位电平从低到高，则表示0
    - 如果中间位电平从高到低，则表示1

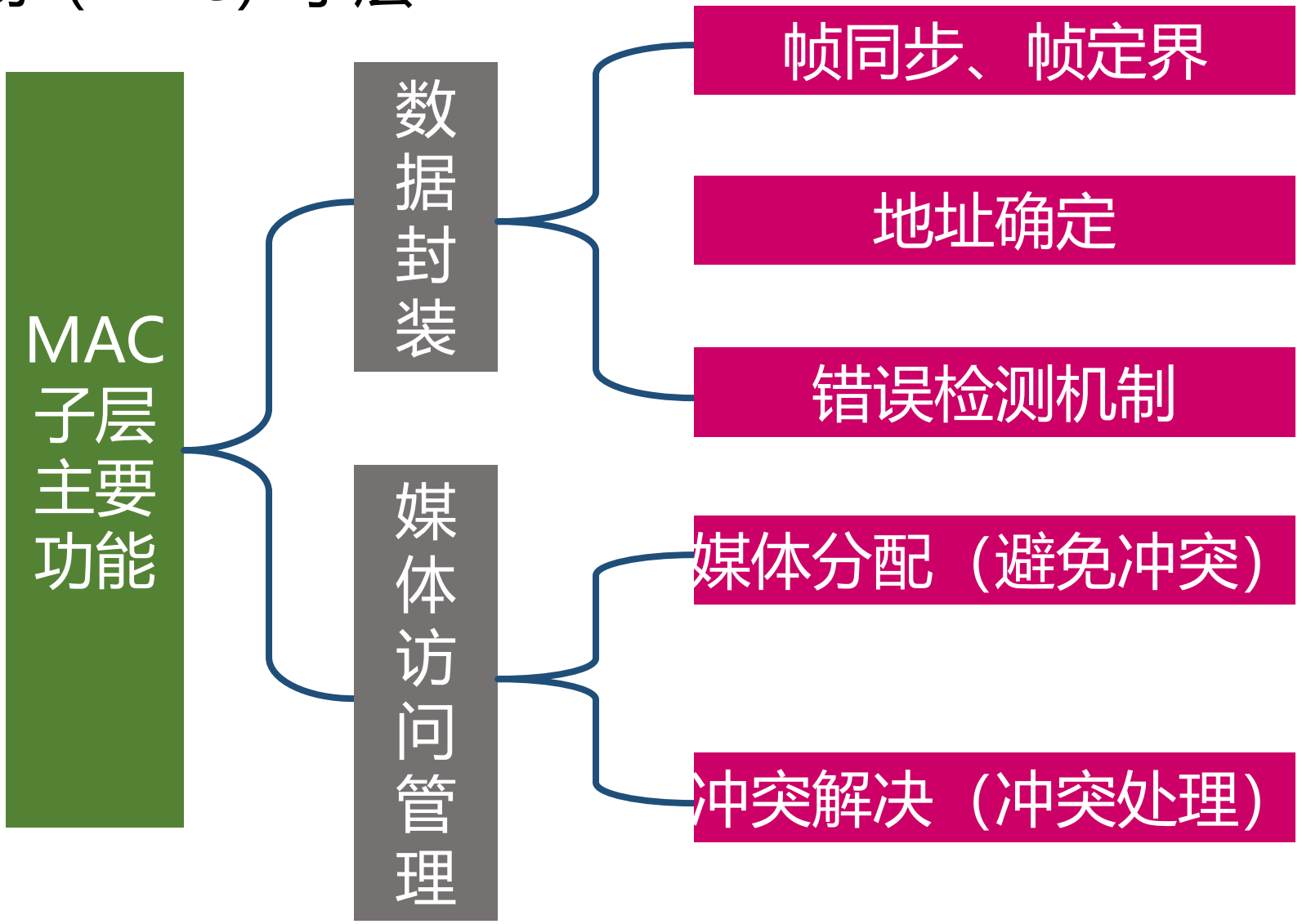
以太网物理层标准

●各种曼彻斯特编码机制示意图 *

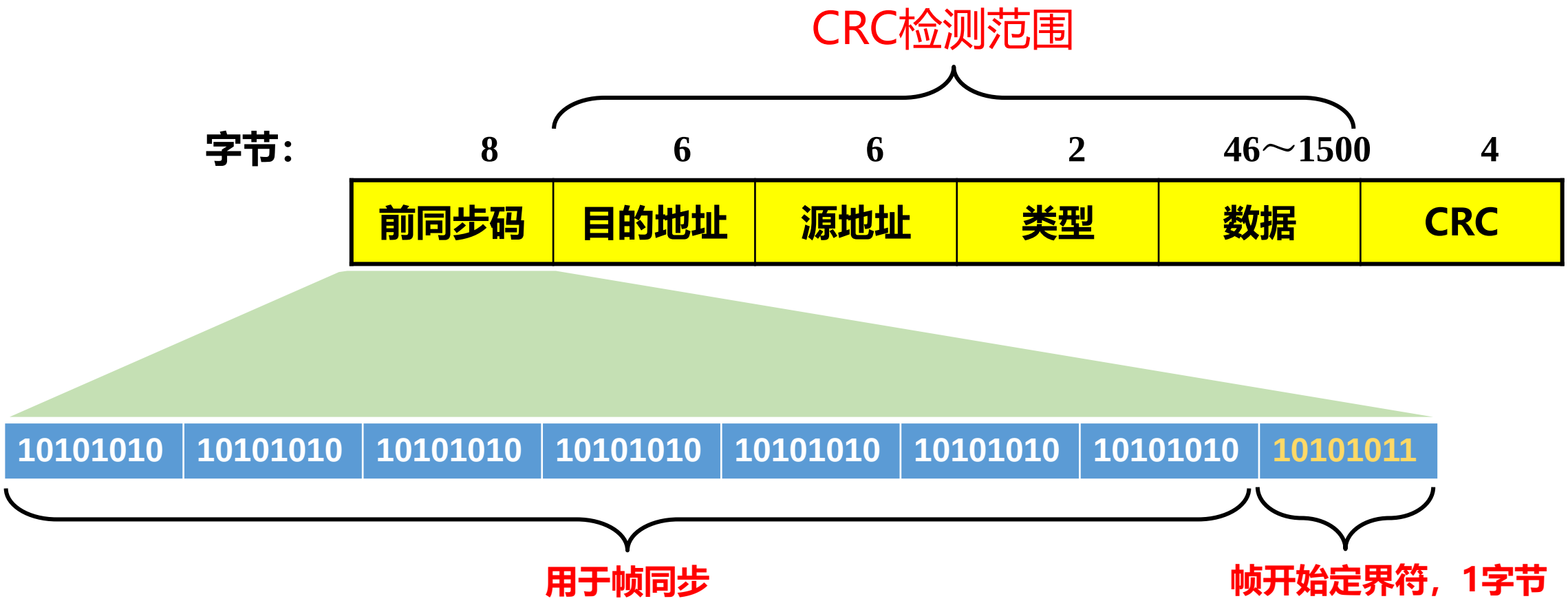


以太网链路层控制技术

- 媒体访问控制（MAC）子层



以太网的帧结构



- 发送方：发送适配器将IP数据报封装成以太网帧，并传递到物理层。
- 接收方：接收适配器从物理层收到该帧，取出IP数据报，并传递给网络层。

前同步码(8 字节)

- 前7字节是“10101010”，最后一个字节是“10101011”。
- 使接收方和发送方的时钟同步，接收方一旦收到连续的8字节前同步码，可确定有帧传过来。
- 前同步码是“无效信号”，接收方收到后删除，不向上层传。
- CRC的校验范围不包括前同步码。

前同步码	目的地址	源地址	类型	数据	CRC

源、目的MAC地址(各6字节)

- 例，同一以太网LAN中两台主机通信。
- 主机A向主机B发送一个IP数据报。

✓主机A适配器的MAC地址：XX-XX-XX-XX-XX-XX

✓主机B适配器的MAC地址：YY-YY-YY-YY-YY-YY

- 适配器B只接收目的地址与其MAC地址匹配或广播地址的帧，并将数据字段的内容传递给网络层。否则，丢弃该帧。

前同步码	目的地址	源地址	类型	数据	CRC

类型字段(2 字节)

以太网可以“多路复用”（支持）多种网络层协议。通过“类型”字段区分。

- 发送方填入网络层协议“类型”编号；
- 接收适配器根据“类型”字段，将数据字段传递给相应的网络层协议。

前同步码	目的地址	源地址	类型	数据	CRC

数据字段(46～1500 字节)

携带网络层传来的IP数据报

- 以太网的最大传输单元MTU是1500字节：
  - 若IP数据报超过1500字节，必须将该数据报分段。
- 最小长度是46字节：
  - 如果IP数据报小于46字节，必须填充为46字节。接收方网络层去除填充内容。

前同步码	目的地址	源地址	类型	数据	CRC

循环冗余检测CRC(4字节)

检测数据帧中是否出现比特差错（翻转）。

✓发送主机计算CRC：范围包括目的地址、源地址、类型、数据字段的比特，结果放入帧CRC字段。

✓接收主机进行CRC校验：接收主机对收到的帧进行同样计算，并校验结果是否和CRC字段的内容相等。若计算结果不等于CRC字段的值(CRC校验失败)，该帧有差错。

前同步码	目的地址	源地址	类型	数据	CRC

以太网: 不可靠的无连接服务

以太网向网络层提供的服务。

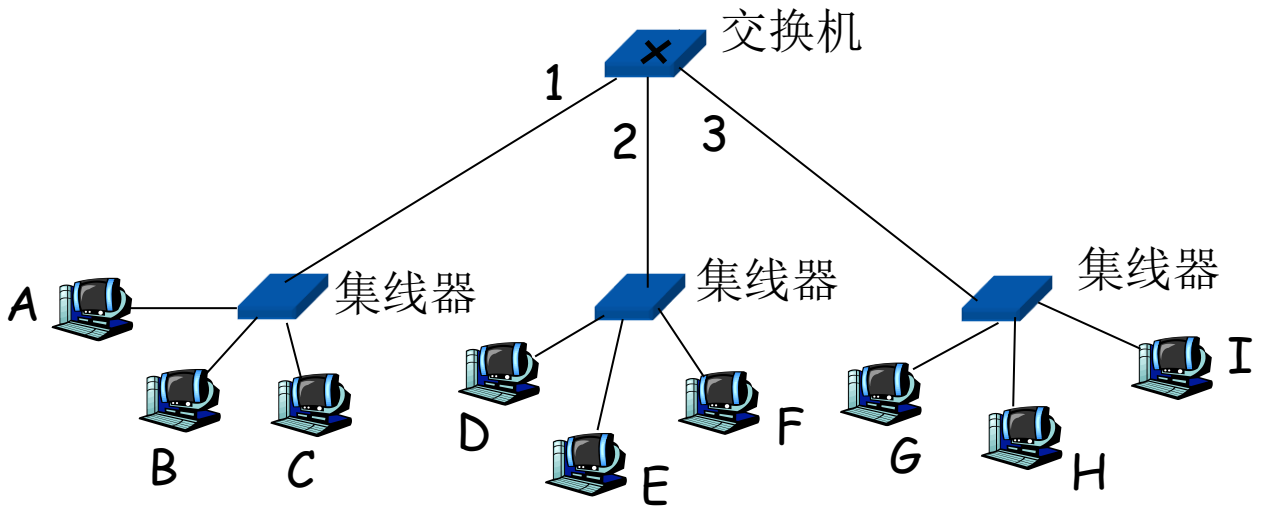
- 无连接服务：通信时，发送方适配器不需要先和接收方适配器“握手”。
- 不可靠的服务：接收到的帧可能包含比特差错。
  - 收到正确帧，不发确认帧；
  - 收到出错帧，丢弃该帧，不发否定帧。
  - 发送适配器不会重发出错帧。
  - 丢弃数据的恢复是通过终端传输层的可靠数据传输机制来实现的
- 以太网的MAC协议：使用无时隙的CSMA/CD协议（二进制指数回退）

  ——见前述5.3.2节

5.4.3 链路层交换机

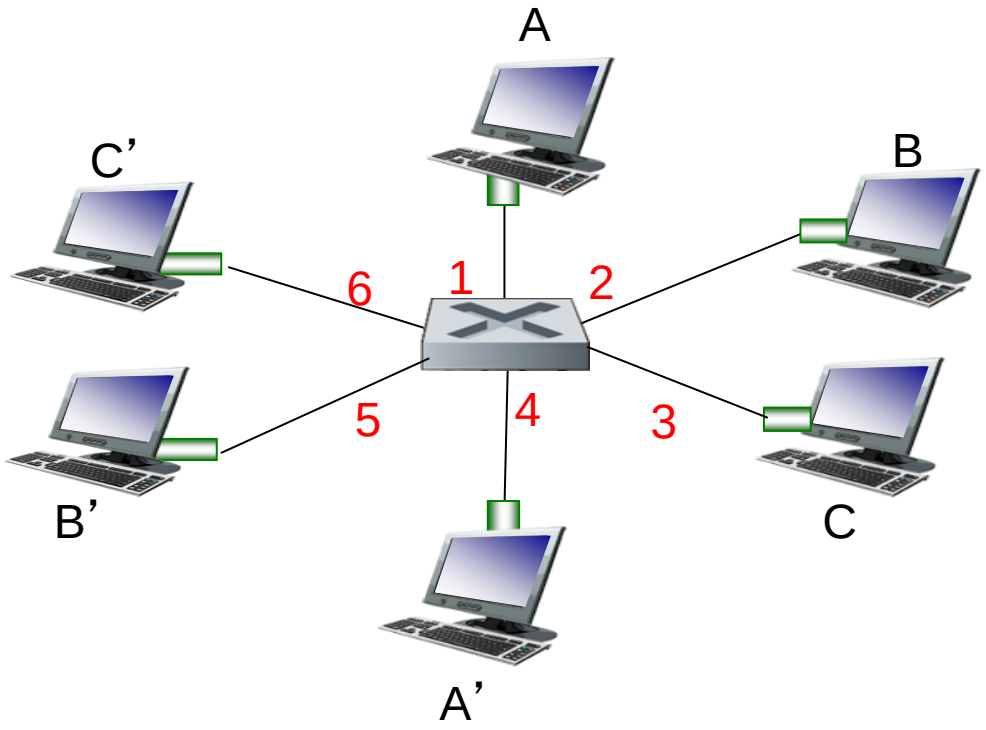
链路层交换机

- 链路层设备
  - 存储转发数据帧
  - 检查达到的数据帧的MAC地址，有选择的转发数据帧到一个或多个输出链路，当数据帧被转发到一个共享网段时，使用CSMA/CD来访问共享链路
- 透明
  - 主机不关心是否存在交换机
- 即插即用和自学习
  - 交换机不需要手工配置



交换机：支持多节点同时传输

- 每个主机由单独的链路直接连到交换机端口
- 交换机可以缓存数据帧
- 以太网协议在每个输入链路使用，无碰撞，全双工
  - 每条链路自身是一个碰撞域
- 交换机： A-to-A ‘和B-to-B’ 可以同时传输，而不会发生碰撞



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

交换机转发表

问题: 交换机是怎么知道A’ 可通过端口4达到， B’ 可通过端口5到达呢?

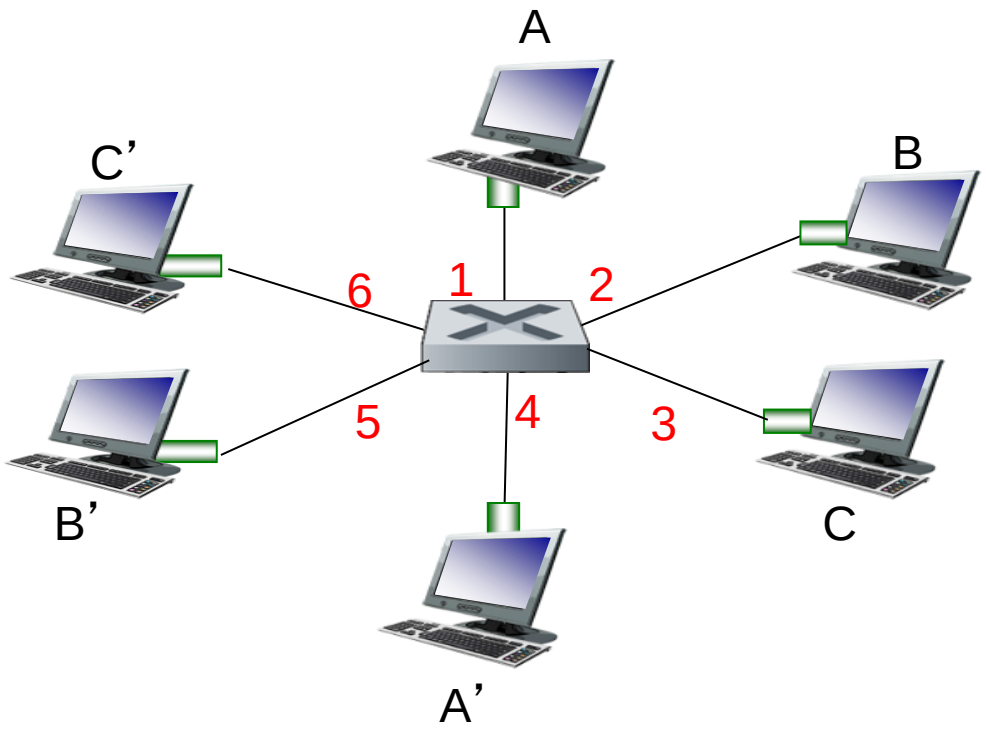
回答：每个交换机有一个交换机转发表，其中每个条目：

(主机的MAC地址，到达主机的端口，时戳)

类似于路由表

问题：转发表中的条目是怎么建立的呢?是否类似于路由协议呢?

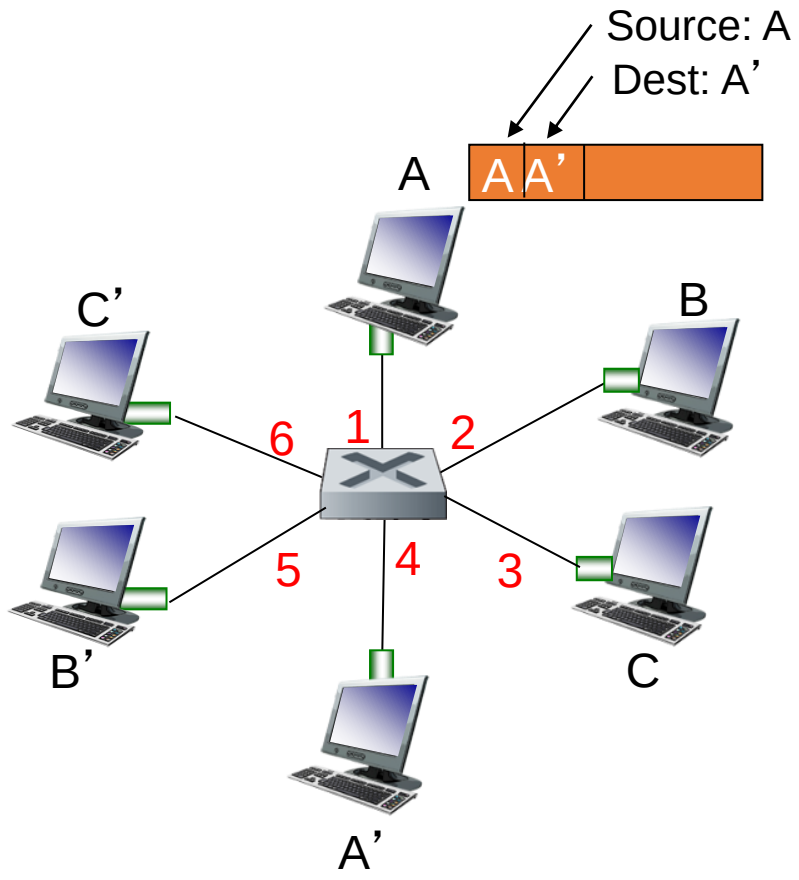
回答：通过自学习



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

交互机：自学习

- 交换机会学习通过哪些端口可以到达哪些主机
  - 当收到数据帧时，交换机“学习”发送主机的位置：进入的局域网网段(到达端口)
  - 在转发表中记录发送主机/位置对



MAC addr	interface	TTL
A	*1*	*60*

Switch table (initially empty)

交换机：数据帧的过滤/转发

当交换机收到数据帧:

1. 记录到达链路和发送主机的MAC地址
2. 使用数据帧的目的MAC地址，在转发表中检索
3. 如果在转发表条目中找到对应的MAC地址
4. 执行{

   如果 目的MAC地址对应的端口与数据帧的达到端口相同

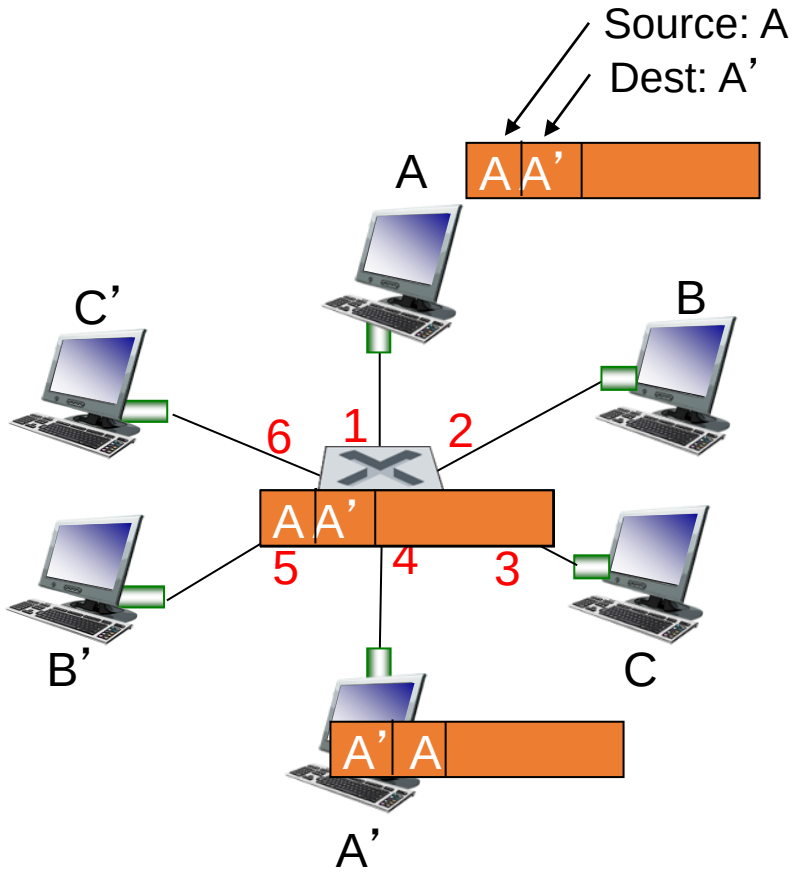
   则 丢弃该数据帧

   否则 转发该数据帧到条目指定的端口
5. }
6. 否则，向除到达端口之外的所有端口转发(flood)

自学习/转发的例子

- 以太帧的目的，A’, 位置未知: *flood*
- 目的A的位置已知:

  在对应的端口1上转发

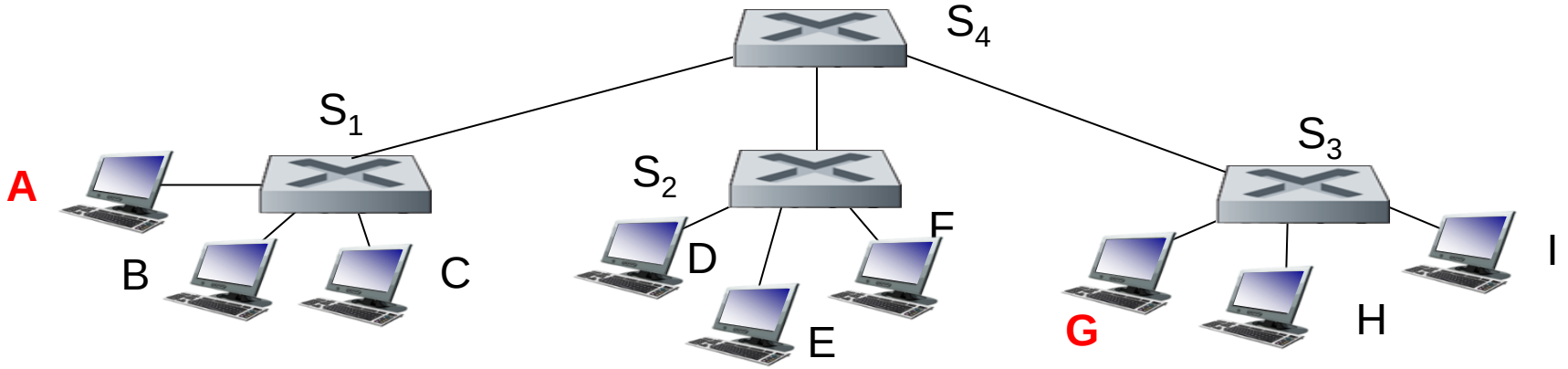


MAC addr	interface	TTL
A	*1*	*60*
A’	*4*	*60*

switch table (initially empty)

交换机互连

- 交换机可以互连在一起

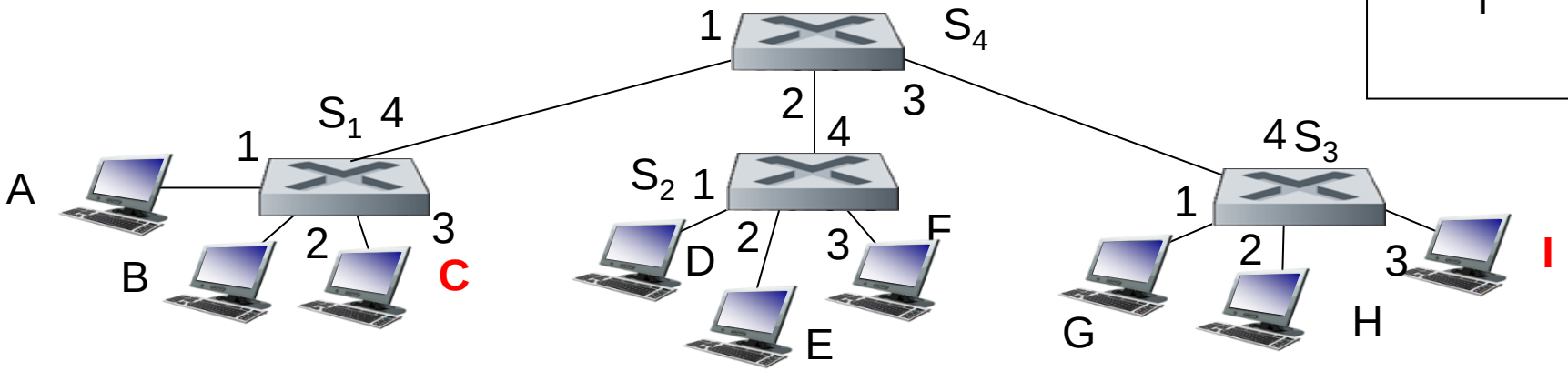


问题：A发送数据帧给G——S1是怎么知道要把数据帧先转发到S4和S3的?

回答：泛洪和自学习

多个交换机自学习的例子

假设主机C发送数据帧到主机I，主机I响应给主机C



S_4

MAC addr	interface	TTL
C	*1*	*60*
I	*3*	*60*

问题: 请大家画出S_1, S_2, S_3, S_4交换机转发表和分组转发

S_1

MAC addr	interface	TTL
C	*3*	*60*
I	*4*	*60*

S_2

MAC addr	interface	TTL
C	*4*	*60*

S_3

MAC addr	interface	TTL
C	*4*	*60*
I	*3*	*60*

交换机的交换特点

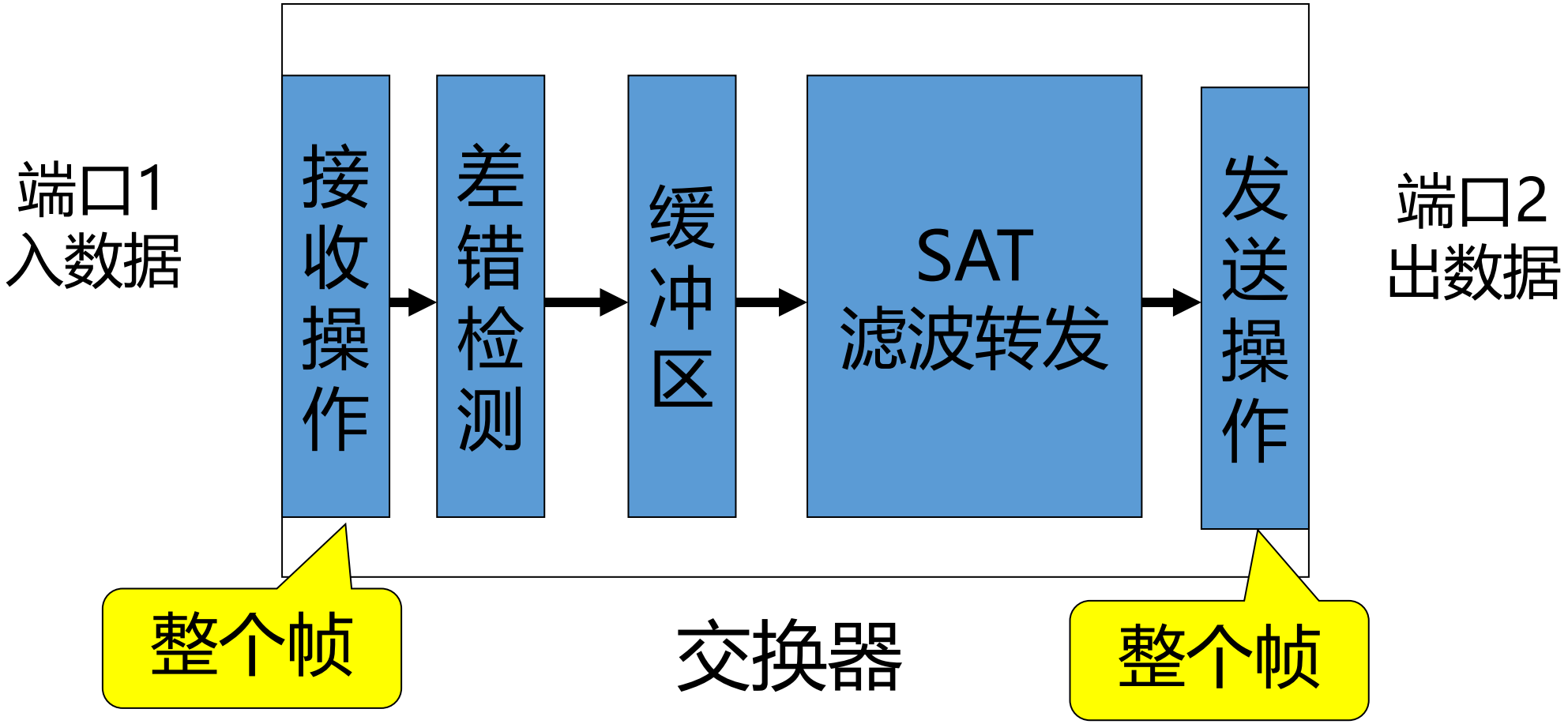
- 识别目的MAC地址，根据交换表进行端口选择
- 识别源MAC地址更新交换表

在识别目的MAC地址和源MAC地址的过程中是否需要接收并缓存完整的帧呢?

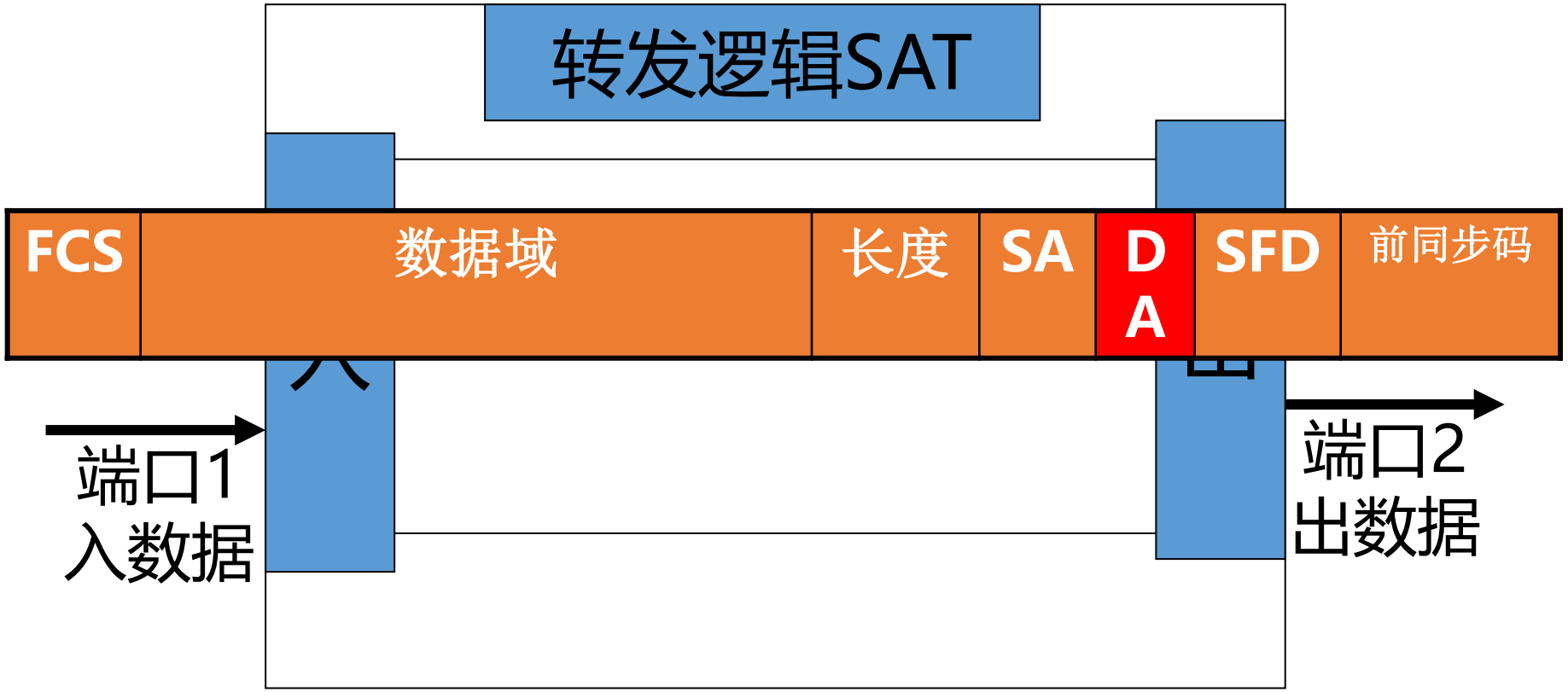
交换机的交换方式

- 存储转发（缓存整个帧后再转发）
- 快速分组又称直通交换（识别出目的地址直接转发）

存储转发交换方式



快速分组交换方式



交换器

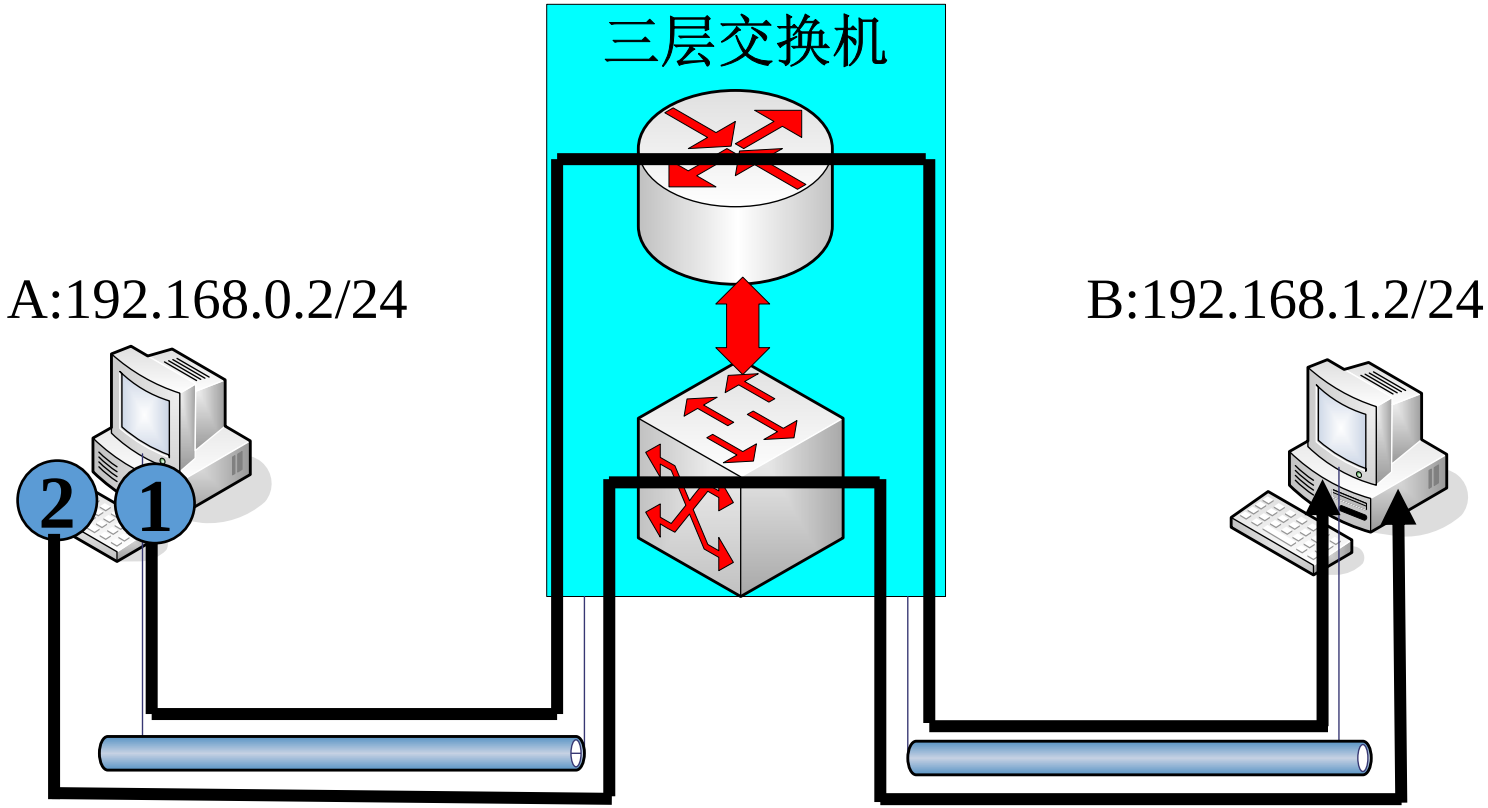
交换机的交换方式讨论

- 存储转发：具有差错检测功能，转发时延较大，适用于出错率高的链路。
- 快速分组又称直通交换：不具有差错检测功能，转发时延较小，适用于时延要求高，出错率低的链路。

三层交换机

- 三层交换是相对于传统的交换概念而提出的
- 传统的交换技术是在OSI网络参考模型中的第二层（即数据链路层）进行操作的，通常称做“二层交换机”。
- 三层交换技术能够在网络模型中的第三层实现数据包的高速转发。
- 简单地说，三层交换技术就是二层交换技术+三层转发技术，三层交换机就是“二层交换机+基于硬件的路由器”

三层交换机的交换机制

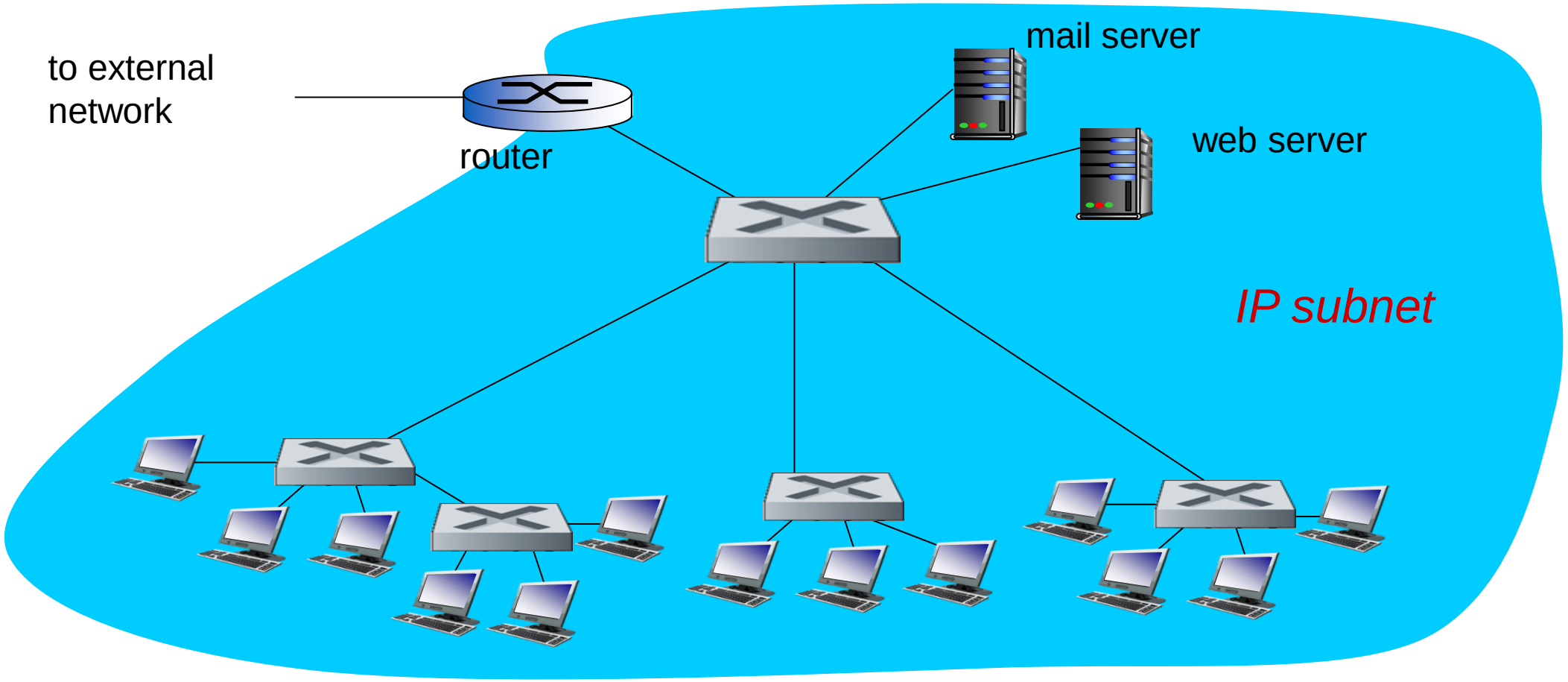


- 一次路由，多次交换

三层交换机的工作原理

- 发送站点A在开始发送时，把自己的IP地址与B站的IP地址比较，判断B站是否与自己在同一子网内。
- 若目的站B与发送站A在同一子网内，则进行二层的转发。
- 若两个站点不在同一子网内，则发送站A要向“缺省网关”发出ARP请求，请求获得B的MAC地址。
- 如果三层交换机知道B的MAC地址，则向A回复B的MAC地址。否则三层交换机根据路由信息向B站广播一个ARP请求，B站得到此ARP请求后向三层交换机回复其MAC地址，三层交换机将B站的MAC地址保存到二层交换引擎的MAC地址表中，并回复给发送站A。
- A直接用B的MAC地址封装数据帧，三层交换机接收到数据后直接进行二层交换。

机构的网络


to external
network
router
mail server
web server
IP subnet

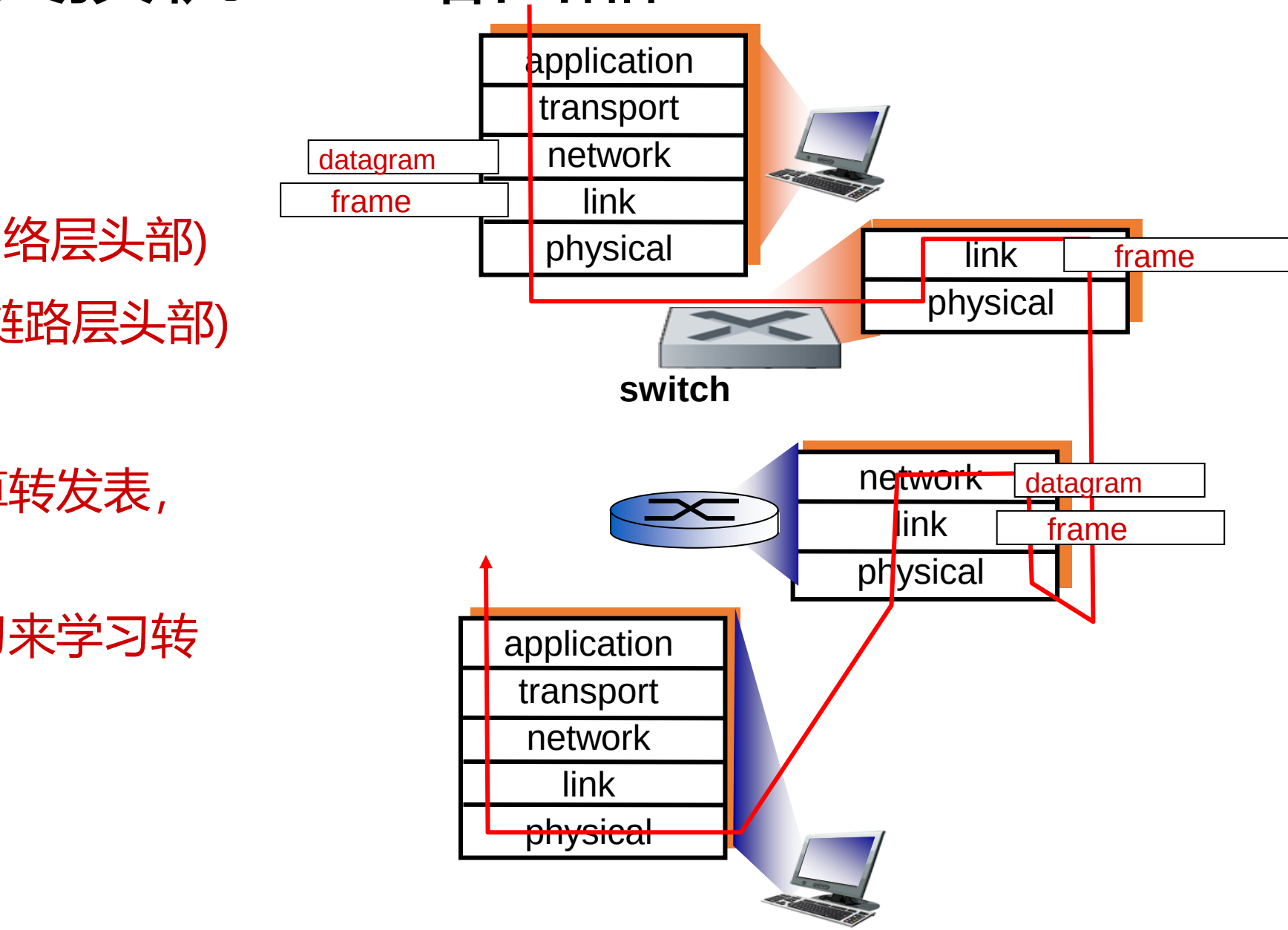
交换机 vs. 路由器

两者都是存储转发设备:

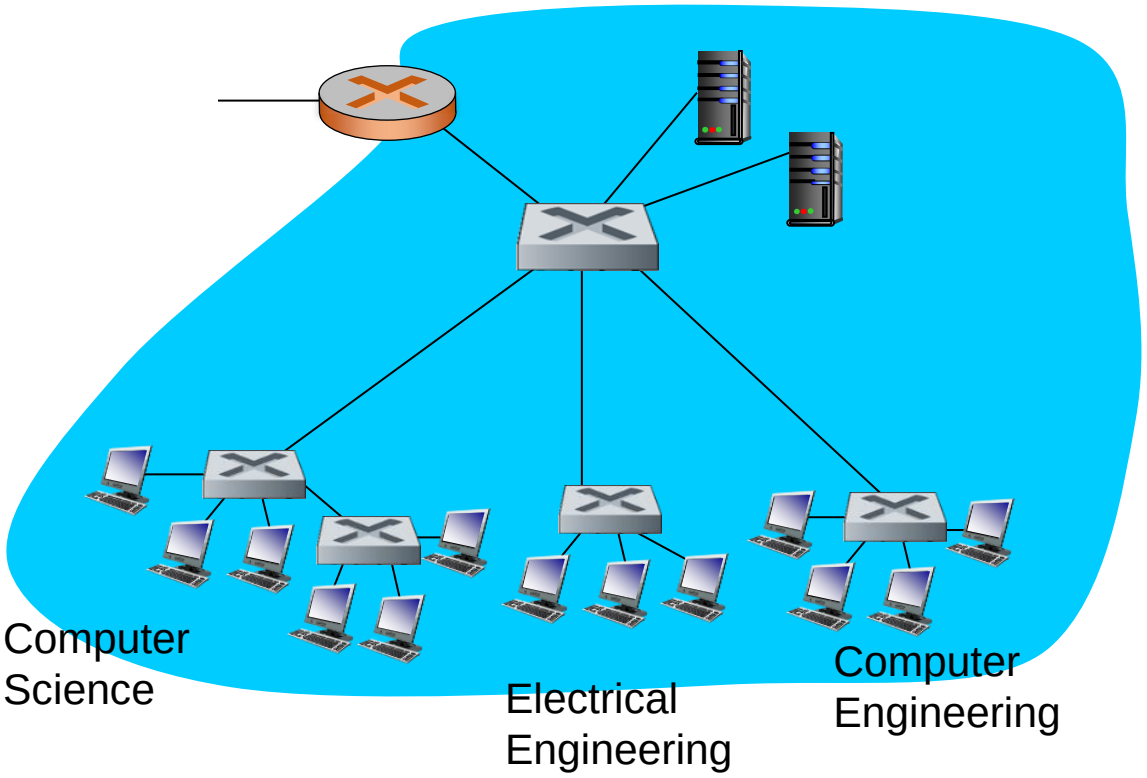
- 路由器: 网络层设备(检查网络层头部)
- 交换机：链路层设备(检查链路层头部)

两者都有转发表

- 路由器：使用路由算法计算转发表，基于IP地址转发
- 交换机：通过泛洪、自学习来学习转发表，基于MAC地址转发



VLAN: 动机



假设：

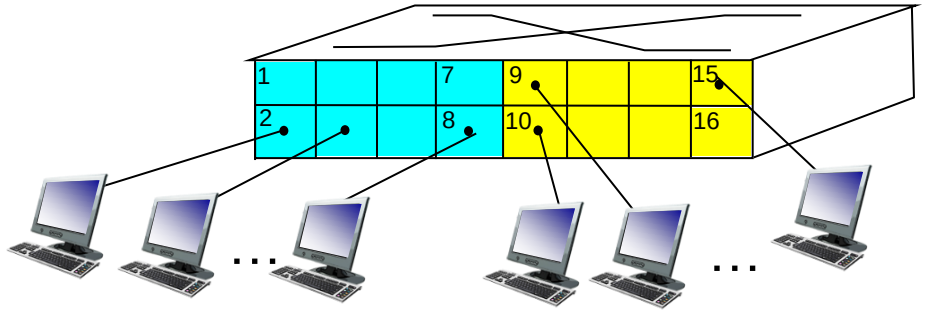
- 计科系(CS)一个用户A到电子工程系办事(EE)，但他想连入到CS的交换机：
- 如果将图中所有设备划分为一个LAN，可以满足A的需求，但这样一个单一的广播域会带来如下问题：
  - 所有的2层广播流量（ARP, DHCP等）会跨越整个网络
  - 安全、隐私、效率等

VLANs

Virtual Local Area Network

利用支持VLAN的交换机，可以在一个实际的物理局域网内，定义多个虚拟的局域网

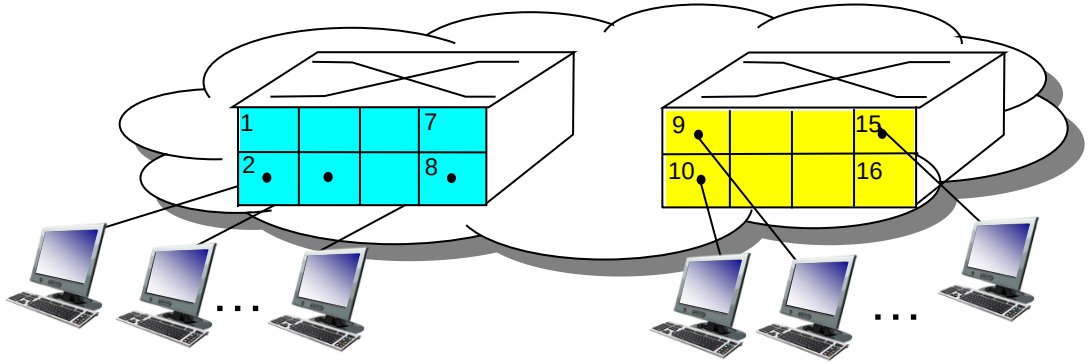
基于端口的VLAN: 利用交换机内置的管理软件，将端口分组，使得一个单独的交换机 ……



Electrical Engineering
(VLAN ports 1-8)

Computer Science
(VLAN ports 9-15)

… 像多个交换机那样工作

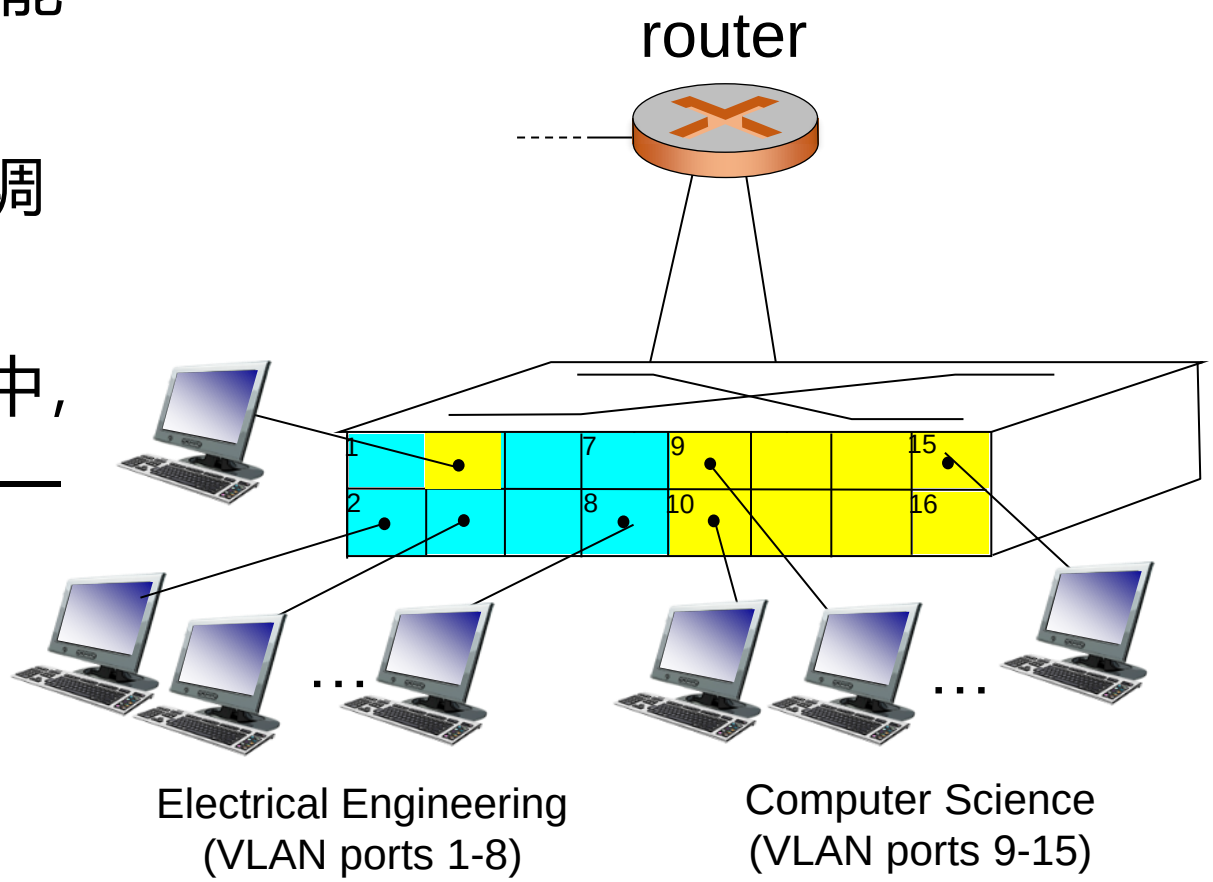


Electrical Engineering
(VLAN ports 1-8)

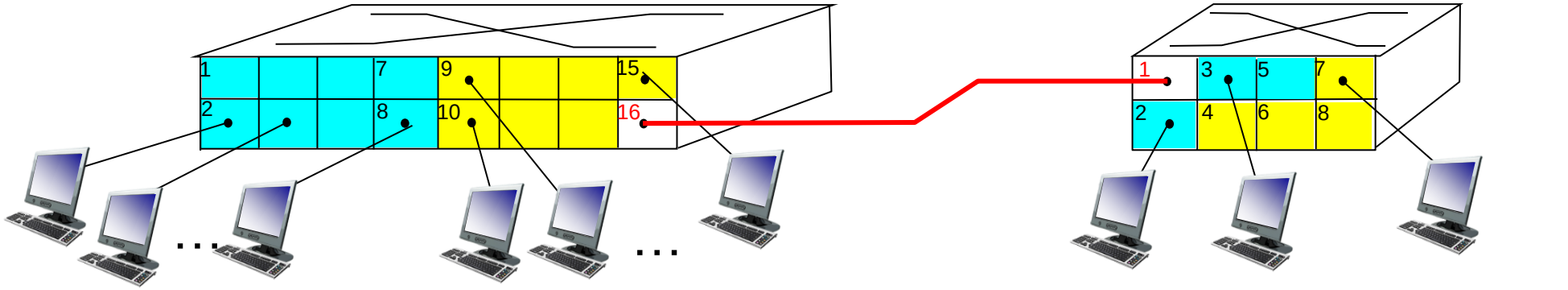
Computer Science
(VLAN ports 9-16)

基于端口的 VLAN

- 流量隔离：从1-8号端口进/出的帧，只能访问1-8号端口
- 动态成员：端口可以在VLAN之间动态调整
- VLAN间转发：通过路由完成（在实际中，厂商会将路由功能和交换功能都整合在一台设备中）



跨越多个交换机的 VLAN



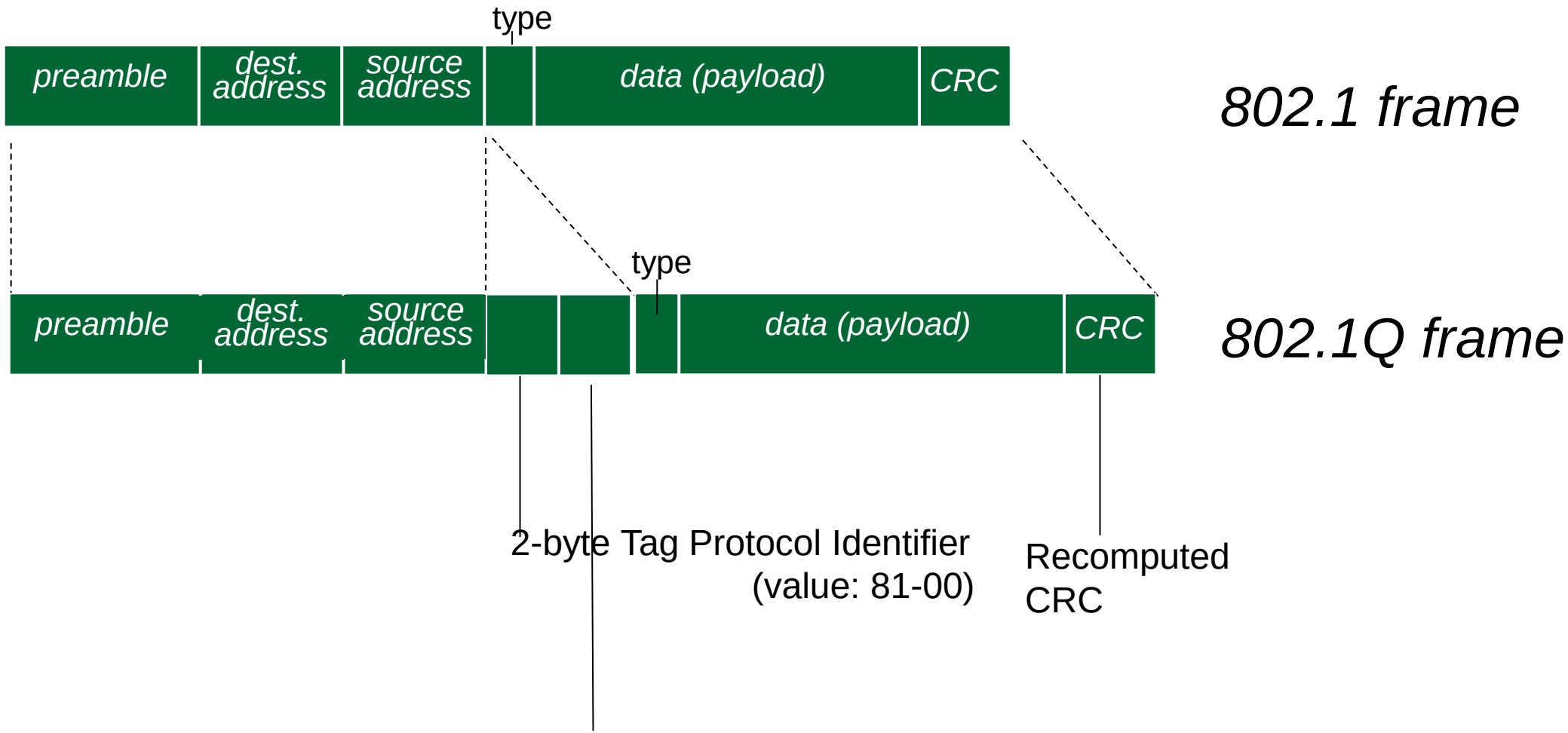
Electrical Engineering
(VLAN ports 1-8)

Computer Science
(VLAN ports 9-15)

Ports 2,3,5 belong to EE VLAN
Ports 4,6,7,8 belong to CS VLAN

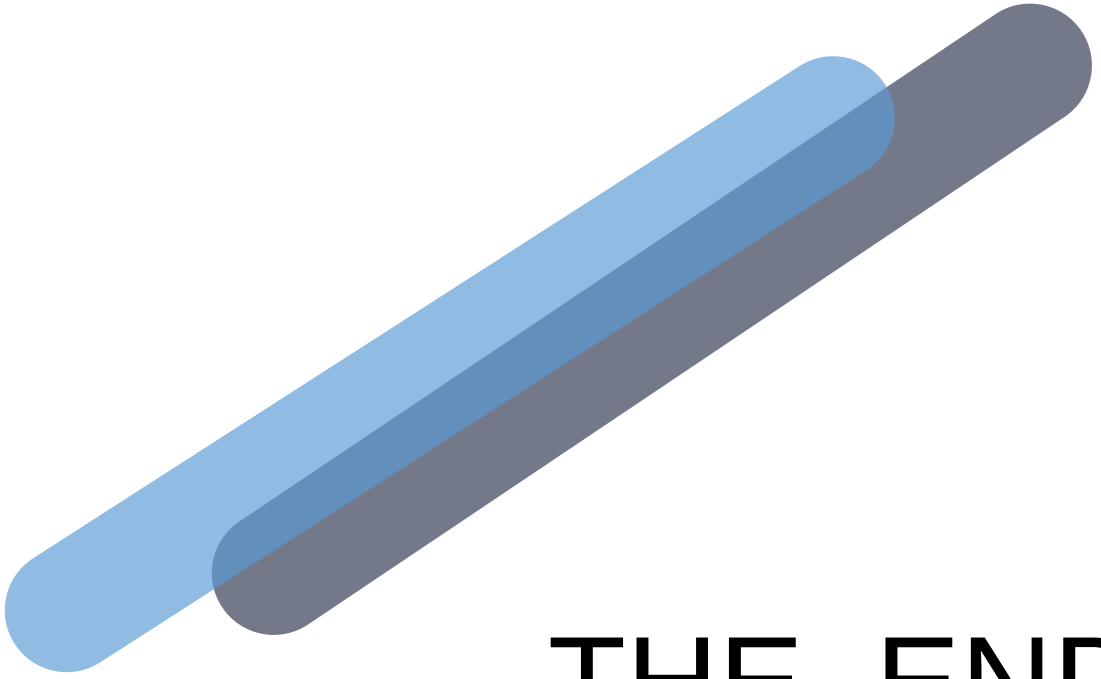
- 干线端口（*trunk port*）承载定义在多个物理交换机之上的VLAN间的流量
- 某一个VLAN内的流量帧，如果要跨域物理的交换机，需使用802.1q格式（带有VLAN ID 信息）
- 802.1q协议的作用：对干线端口之间传输的帧，添加/移除额外的头部字段

802.1Q VLAN 帧格式*



第五章: 总结

- 链路层功能、提供的服务
- 差错检测方法
- 多路访问链路和协议
- 局域网技术
- 以太网技术(链路层和物理层的实现方式)
  - 帧格式
  - 以太网提供无连接、不可靠的服务
  - 以太网采用的CSMA/CD原理
  - 物理层曼彻斯特编码*
- 集线器和交换机
- VLAN



THE END