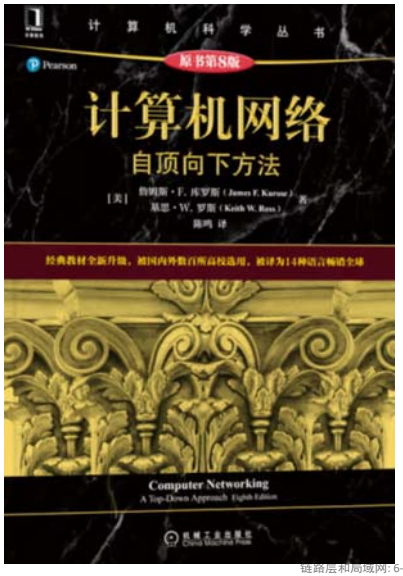
Chapter 6
链路层和局域网
The Link Layer and LANs



第六章：链路层和局域网

章节目标：

- 理解链路层服务背后的原理：
  - 差错检测和纠正技术
  - 共享广播信道：多路访问
  - 链路层寻址
  - 局域网：Ethernet, VLANs
- 数据中心网络
- 各种链路层技术的实例化与实现

链路层和局域网：路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

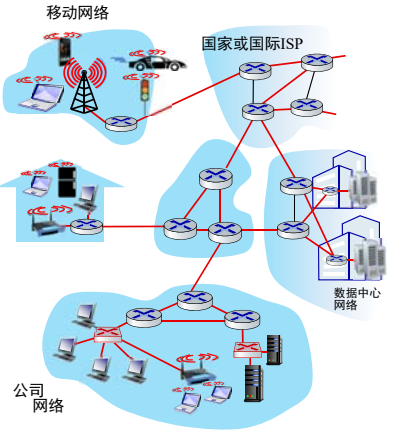
- Web页面请求的历程

链路层概述

术语：

- 节点：运行链路层协议的任何设备
  - 如主机和路由器
- 链路：沿着通信路径连接相邻节点的通信信道
  - 有线的，如以太网链路
  - 无线的，如wifi链路
- 链路层*帧*：layer-2,封装了数据报

*链路层*负责将数据报从一个节点传输到链路上*物理相邻*的另一个节点



链路层概述

- 不同的链路用不同的链路层协议来传输数据报：
  - 例如，第一条链路是WiFi，下一条链路可以是Ethernet
- 每种链路层协议提供不同的服务
  - 例如，是否提供可靠的链路传输服务

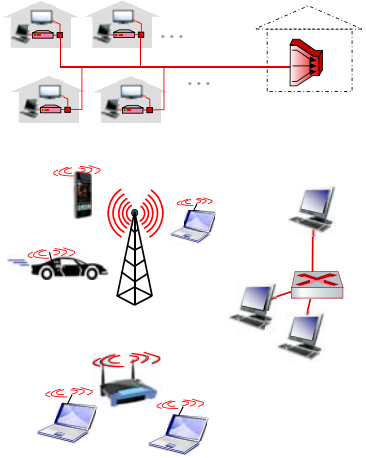
交通运输的类比：

- 一旅行社计划为游客开辟从美国普林斯顿到瑞士洛桑的旅游线路，其认为对游客而言最便利的方案
  - 豪华大轿车：普林斯顿到JFK机场
  - 飞机：JFK机场到日内瓦机场
  - 火车：日内瓦机场到洛桑火车站
- 一个游客 = 一个数据报
- 每个运输区段 = 一条通信链路
- 每种运输方式 =一种 链路层协议
- 旅行社 = 路由选择协议

3段中的每一段的两个相邻地点之间的直达的

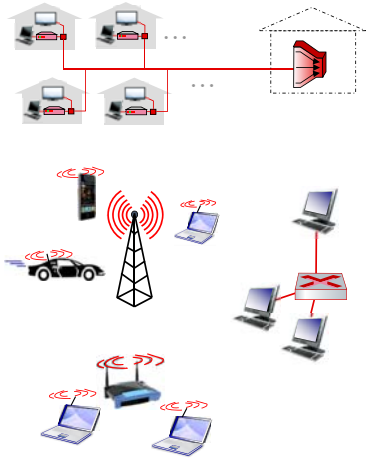
链路层提供的服务

- 成帧
  - 成帧：将数据报封装到帧中，添加首部、尾部
- 链路接入
  - 链路接入: 媒体访问控制协议（MAC协议）规定了帧在链路上传输的规则
  - 帧报头中的“MAC”地址标识源和目的(和IP地址不同！)
- 可靠交付（相邻节点之间的）
  - 与TCP类似，链路层可靠交付服务通常是通过确认和重传取得
  - 很少在低比特差错率的链路上使用，因为被认为是不必要的开销
  - 无线链路：高比特差错率
    - Q: 为什么链路层和端到端都有可靠交付?
    - A: 其目的是本地（即在差错发生的链路上）纠正一个差错，而不必迫使进行端到端的数据重传



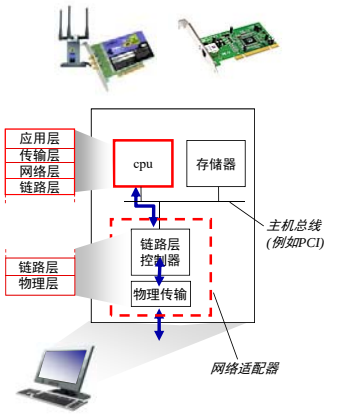
链路层提供的服务 (续)

- 差错检测和纠正:
  - 差错检测：
    - 信号衰减引起的差错, 噪音.
    - 接收节点检测到差错, 信号重传, 或丢弃帧
    - 用硬件实现
  - 差错纠正:
    - 接收节点识别并*纠正*比特差错而不需要重传
- 链路层提供的其他服务
  - 流量控制:
    - 相邻的发送和接收节点的步调同步
  - 半双工和全双工:
    - 在半双工的情况下，链路两端的节点可以传输数据，但不能同时传输



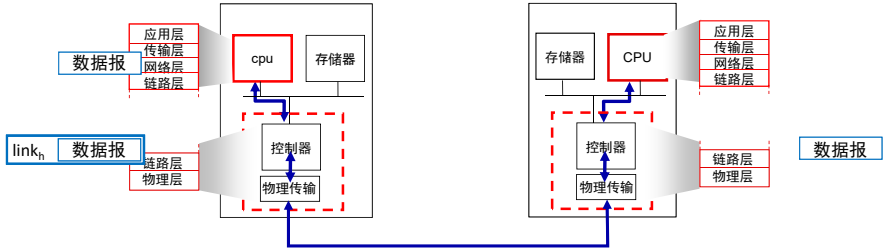
链路层在何处实现?

- 每一台主机中
- 链路层实现在*网络适配器*(也称为*网络接口卡NIC*) 或芯片上
  - 链路层的许多功能（成帧、链路接入、差错检测等）是由硬件实现的
    - 如Ethernet网卡、 WiFi网卡或者芯片
  - 90年代以前，大部分网络适配器是物理上分离的卡（如PCMCIA卡）;后期集成在主板上
  - NIC或者芯片实现了链路层和物理层
- 连接到主机的系统总线上
- 硬件和软件的结合体，即此处是协议栈软件和硬件交接的地方
  - 大部分链路层功能由硬件实现，部分链路层是由运行于主机CPU上的软件中实现的（如组装链路层寻址信息和激活控制器）



网卡

接口通信



发送端控制器：

- 取得由协议栈较高层生成并存储在主机内存中的数据报
- 在帧中封装数据报
- 增加差错检测比特、可靠数据传输、流量控制等将该帧传进通信链路

接收端控制器：

- 查找差错、可靠数据传输、流量控制等链路层服务
- 链路层软件响应控制器中断，处理差错条件，提取数据报，传递到接收方的上层

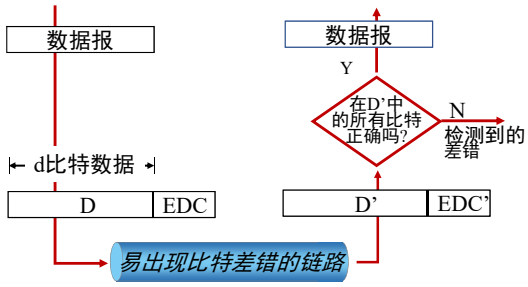
链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

差错检测

使用EDC（差错检测和纠正比特 Error Detection and Correction）来增强数据D，(如冗余)
要保护的数据：来自网络层的数据报+链路层首部字段



*D:要保护的数据

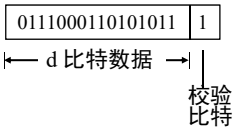
差错检测不是100%可靠！

- 协议可能会漏掉一些错误，但很少发生
- EDC字段长度越大，则检测和纠正效果越好

奇偶校验

单比特奇偶校验:

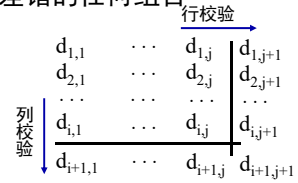
- 检测单比特差错



偶校验: 发送方只需包含一个附加的比特，设置它的值，使得d+1比特中1的个数为偶数

二维奇偶校验:

- 对D中的d个比特被划分为i行j列，产生i+j+1差错检测比特
- 检测*和纠正* 单比特差错，可检测（但不能纠正）2比特差错的任何组合



无差错:

1	0	1	0	1	1
1	1	1	1	0	0
0	1	1	1	0	1
1	0	1	0	1	0

可纠正的单比特差错:

1	0	1	0	1	1
1	0	1	1	0	0
0	1	1	1	0	1
1	0	1	0	1	0

校验错 (行)

校验错 (列)

* Check out the online interactive exercises for more examples: http://gaia.cs.umass.edu/kurose_ross/interactive/

检验和方法

回顾下UDP检验和（第三章）:在传输的报文段中检测“差错”（如比特翻转）-第三章

发送方:

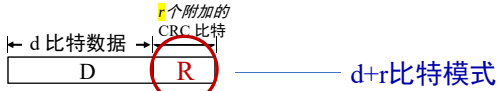
- 将报文段内容处理为16比特整数序列
- 检验和 checksum: 报文段内容的加法（反码和）
- 发送方将检验和放入UDP检验和字段

接收方:

- 计算接收的报文段的检验和
- 核对计算的检验和是否等于检验和字段的值:
  - 不相等 - 检测到差错
  - 相等 – 未检测到差错。虽然如此，但是否可能会有差错?

循环冗余检测(CRC，Cyclic Redundancy Check)

- CRC也称为多项式编码
- 更强的差错检测编码，是链路层常用的差错检测方法，由专用的硬件实现
  - TCP，UDP差错检测使用软件实现，通常采用简单而快速如检验和这样的差错检测
- D: 被发送的数据，*d* 比特数据 (可看成一个二进制数)
- G: 生成多项式 (给定的、发送方和接收方协商好的),也叫作*r+1*比特模式，最高位和最低位必须均为1



目标: 选择 *r* 个 附加比特, R使得 <D,R> 能够被 G (模 2 算术) 整除

- 接收方知道 G, 把 <D,R> 除以 G. 如果得到非零余数: 检测到差错!

循环冗余校验 (CRC): 例子

要求:

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

等价于:

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

等价于:

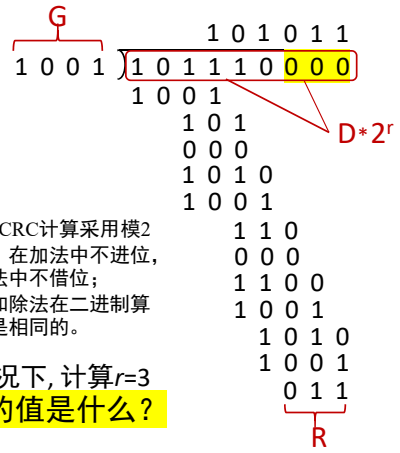
如果我们用 $D{\cdot}2^r$ 除以 G, 余数 R :

$$R = remainder\left[\frac{D{\cdot}2^r}{G}\right]$$

例子:

D=101110, 即6比特数据, 在G=1001情况下, 计算*r*=3

计算：D=1001010101, G=10011, R的值是什么？

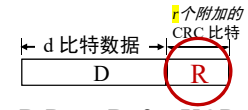


*所有CRC计算采用模2算术，在加法中不进位，在减法中不借位；乘法和除法在二进制算术中是相同的。

* Check out the online interactive exercises for more examples: http://gaia.cs.umass.edu/kurose_ross/interactive/

循环冗余检测(CRC，Cyclic Redundancy Check)

- D: 被发送的数据，*d* 比特数据 (可看成一个二进制数)
- G: 生成多项式 (给定的、发送方和接收方协商好的),也叫作*r+1*比特模式，最高位和最低位必须均为1



d+r比特模式---发送方发送的数据

$<D,R> = D*2^r \ \text{XOR} \ R$ ——— <D,R>数学公式满足： $D\,2^r \ \text{XOR} \ R = nG$

用CRC进行差错检测:

接收方用G去除收到的接收到的d+r比特。
若余数为0，则数据正确，接收数据；
如果得到非零余数: 检测到差错!

- 可以检测所有小于或等于 *r* 比特的突发差错；在适当的假设下，长度*r+1*比特的差错以$1-0.5^r$概率被检测到
- 被广泛应用到实践中 (Ethernet, 802.11 WiFi)

链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

多路访问链路和协议

两种类型的网络“链路”：

- 点对点链路：由链路一端的单个发送方和链路另一端的单个接收方组成
  - 以太网交换机与主机之间的点对点链路
  - 点对点协议(PPP)用于拨号访问，高级数据链路控制协议(HDLC)用于可靠的任何一种比特流的点对点传输
- 广播链路（共享线缆或物理媒体）：让多个发送和接收节点连接到相同的、单一的、共享的广播信道上
  - 传统 Ethernet
  - 有线接入网的上行HFC(混合光纤同轴电缆)
  - 802.11 无线 LAN, 4G/5G, 卫星


有线共享(例如以太网电缆)


无线共享：4G/5G


无线共享：WiFi


无线共享：卫星


鸡尾酒派对上的人们
共享空气，声音

多路访问协议

- 单共享广播信道
- 节点同时进行两次或两次以上的传输：干扰
  - 如果节点同时接收到两个或多个信号就发生*碰撞*

多路访问协议(MAC, Multiple Access Protocol)

- 确定节点如何共享信道的分布式算法，如确定节点何时传输
- 信道共享的通信必须使用信道本身！
  - 没有带外信道用于协调

一种理想的多路访问协议

给定: 速率为 *R* bps的广播信道

希望:

1. 当仅有一个节点发送数据时，该节点具有*R* bps的吞吐量
2. 当有M个节点发送数据时，每个节点吞度量为*R/M* bps。这不必要求M个节点中的每个节点总有*R/M*的瞬时速率，而使每个节点在一些适当定义的时间间隔内应该有*R/M*的平均传输速率
3. 协议是分散的:
   - 没有特殊的节点来协调传输，如不会因为某个主节点故障而使整个系统崩溃
   - 没有时钟、时隙的同步
4. 协议是简单的，使实现不昂贵

多路访问协议的分类

三大类:

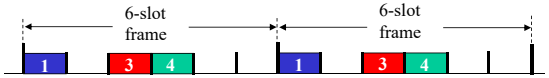
- 信道划分协议
  - 将信道分成更小的“片（piece）”(时隙、频率、编码)
  - 将片分配给节点单独使用
- 随机接入协议
  - 信道不分割，允许碰撞
  - 从碰撞中“恢复”
- 轮流协议
  - 节点轮流发送数据，但可能有更多数据要发送的节点得到更长的时间段来发送其数据

信道划分协议: TDMA

TDMA: 时分多址(time division multiple access)

- “轮流”访问信道
- 将时间划分为时间帧（time frame），并进一步将每个时间帧划分为N个时隙(slot)
- 每个节点在每轮得到固定长度的时隙(一个时隙长度=单个分组的传输时间)
- 未使用的时隙空闲

例: 6个节点的LAN，每个时间帧划分为6个时隙，1,3,4节点有数据包要发送，时隙2,5,6空闲

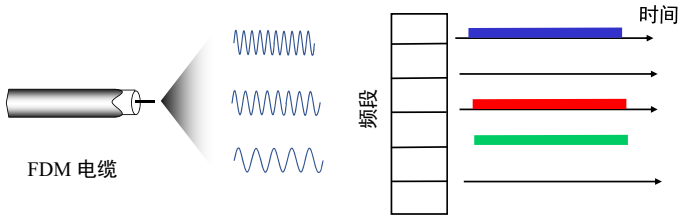


信道划分协议: FDMA

FDMA: 频分多址(frequency division multiple access)

- 信道频谱划分为频段
- 每个节点被分配固定的频段
- 在频段中未使用的传输时间空闲

例: 6个节点的LAN，划分了6个频段，1,3,4节点有数据要发送，频段2,5,6空闲



信道划分协议: TDMA

CDMA: 码分多址 (Code Division Multiple Access)

- 为每个用户分配不同的编码；即编码集划分
- 所有用户使用相同的频率发送数据，但是每个用户使用自己唯一的码片序列（即编码）对它发送的数据进行编码
  - 允许多个用户“同时存在”并且以最小的干扰同时传输
  - 码片序列需要“正交”
- 接收方能够正确接收发送方编码的数据比特，即使有多个发送方同时的干扰传输
  - 编码：内积：（原始数据）*（码片序列）
  - 解码：内积和：∑（编码数据）*（码片序列）

随机接入协议

- 当节点有数据包要发送时
  - 以信道的全部速率 R 传输
  - 节点之间不存在 *事先* 协调
- 两个或多个正在传输节点:“碰撞”
- 随机接入协议主要规定了:
  - 如何检测碰撞
  - 如何从碰撞中恢复 (例如，通过延迟重传)
- 常用的随机接入协议的例子:
  - ALOHA, 时隙 ALOHA
  - CSMA（ Carrier Sense Multiple Access，载波侦听多路访问），CSMA/CD（CSMA with Collision Detection，具有碰撞检测的载波侦听多路访问)

时隙 ALOHA

假定:

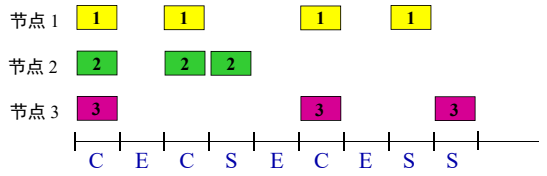
- 所有帧由L比特组成，即大小相同
- 时间被划分为长度为*L/R*秒的时隙，即 传输一帧的时间
- 节点只在时隙起始点开始传输帧
- 节点是同步的，每个节点都知道时隙何时开始
- 如果一个时隙中有2个或2个以上的节点传输帧，则所有节点在该时隙结束之前检测到该碰撞事件

操作:

- 当节点有一个新帧要发送时，它等到下一个时隙开始并在该时隙传输整个帧
  - *如果没有碰撞：*该节点成功地传输它的帧
  - *如果有碰撞：该*节点在时隙结束之前检测到这次碰撞。该节点以概率p在后续的每个时隙重传它的帧，直到该帧被无碰撞地传输出去

随机 – *为什么*?

时隙 ALOHA



C: 碰撞时隙
S: 成功时隙
E: 空闲时隙

优点:

- 单活跃节点可以以全信道传输速度*R*连续传输数据
- 高度分散的:仅是节点需要对时隙同步，其他的都是分散的：每个节点检测碰撞并独立地决定何时重传
- 简单

缺点:

- 碰撞，浪费时隙
- 空闲时隙
- 需要时钟同步

时隙 ALOHA: 效率

成功时隙：刚好有一个节点传输的时隙称为成功时隙

时隙多路访问协议的效率: 当有大量的活跃节点且每个节点总有大量的帧要发送时，长期运行中成功时隙的份额。

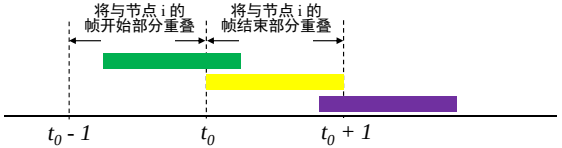
- *假设*: N 个节点每个节点总有帧要发送，每个节点试图在每个时隙内以概率p传输一帧
  - 一个给定节点在时隙内成功传送的概率 = $p(1-p)^{N-1}$
  - N个*节点*、*任意*一个节点成功传送的概率 = $Np(1-p)^{N-1}$
  - 求出使这个表达式最大化的p*
  - **最大效率**: N趋于无穷时，取 $Np^*(1-p^*)^{N-1}$ 的极限，得到：

最大效率 = 1/e =0.37

当有大量节点有很多帧要传输时，则最多仅有37%的时隙做有用的工作。信道的有效传输速率不是Rbps,而仅为0.37Rbps

纯 ALOHA[Abramson 1970]

- 非时隙 Aloha: 更简单、无同步、完全分散的协议
  - 当一帧首次到达时：节点立即将该帧完整地传输进广播信道
- 在没有同步的情况下，碰撞概率会增加：发生碰撞时，每个节点以概率p重传该帧
  - 在 t_0 发送的帧与在 $[t_0-1,t_0+1]$ 发送的其它帧发生碰撞



- 给定节点成功传输一次的概率= $p(1-p)^{2(N-1)}$
- 纯 Aloha 的最大效率**= 1/(2e) =0.18** !

载波侦听多路访问

CSMA (Carrier Sense Multiple Access)

简单的 CSMA: 传输之前先听：

- 如果侦听到信道空闲：传输整个帧
- 如果侦听到信道繁忙：延迟传输，直到检测到一小段时间没有传输，然后再开始传输

- 类比：说话之前先听，如果别人正在说话，等他们说完为止。

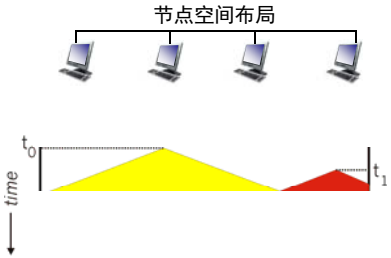
CSMA/CD:具有*碰撞检测*的CSMA

- 当一个节点在传输时一直在侦听此信道
- 检测到碰撞后就停止传输，减少信道浪费
- 重复“侦听-当空闲时传输”循环之前等待一段随机时间
- 在有线传输中容易做到碰撞检测，但在无线传输中难以做到

- 类比：有礼貌的健谈者：如有与他人同时开始说话，则停止说话

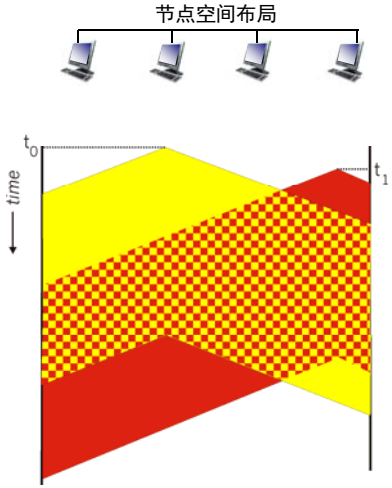
CSMA: 碰撞

- 在载波侦听的情况下仍可能发生碰撞：
  - 信道传播时延意味着两个节点可能无法听到对方刚刚开始的传输
- 碰撞：整个数据包传输时间浪费
  - 距离和传播时延是决定碰撞可能性的重要因素



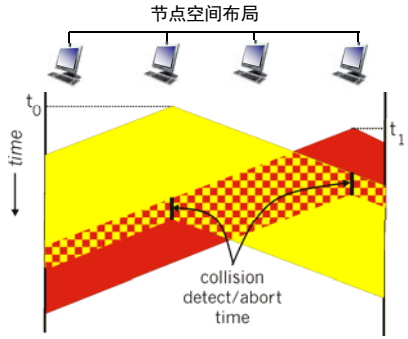
CSMA: 碰撞

- 在载波侦听的情况下仍可能发生碰撞：
  - 信道传播时延意味着两个节点可能无法听到对方刚刚开始的传输
- 碰撞：整个数据包传输时间浪费
  - 距离和传播时延是决定碰撞可能性的重要因素



具有碰撞检测的载波侦听多路访问
CSMA/CD （CSMA with Collision Detection）：

- 减少在碰撞中浪费的时间
  - 在检测到碰撞时传输中止



以太网 CSMA/CD 算法

1. 网络适配器从网络层获得数据报，准备链路层帧，并将其放入网络适配器缓存中
2. 网络适配器侦听信道：
   如果空闲：开始传输帧
   如果忙：等到信道空闲再传输
3. 在传输过程中，网络适配器监听来自其他使用该信道的适配器的信号能量的存在。如果传输整个帧时未检测到来自其他适配器的信号能量，该网络适配器就完成了该帧的传输！
4. 如果网络适配器在传输过程中检测到来自其他适配器的信号能量，它中止传输
5. 中止传输后，网络适配器执行*二进制指数后退*算法：
   - 第 m （m的最大值在10以内）次碰撞后，网络适配器从 $\{0,1,2,3,\ldots,2^m-1\}$ 中等概率地选择 K。网络适配器 等待 $K*512$ 比特时间（即发送512比特进入以太网所需时间的K倍），返回步骤 2
     - 对于100Mbps以太网来说512比特时间为5.12μs
     - 在10次以上的碰撞后，从 $\{0,1,2,3,\ldots,1023\}$ 中等概率选择 k
   - 选择k的集合长度随碰撞次数呈指数增长；更多碰撞：更长的后退间隔

CSMA/CD 效率

- 效率：当有大量的活跃节点且每个节点总有大量的帧要发送时，帧在信道中无碰撞地传输的那部分时间在长期运行时间中所占的份额。
- t_{prop} = 信号能量在任意两个网络适配器之间传播所需要的最大时间
- t_{trans} = 传输一个最大长度的以太网帧的时间

$$\text{效率} = \frac{1}{1 + 5t_{prop}/t_{trans}}$$

- 效率接近于 1
  - 当 t_{prop} 接近 0 （碰撞的节点立即中止而不会浪费信道）
  - 当 t_{trans} 接近无穷（当一个帧取得了信道时，它将占用信道很长时间；因此信道在大多数时间会有效的工作）
- 性能优于 ALOHA: 简单，廉价，分散！

轮流（Taking Turns）协议

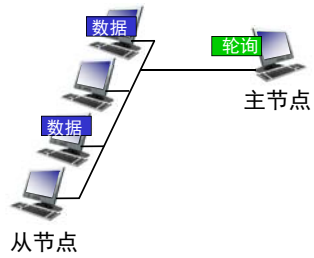
信道划分协议:
- 高负载时*有效*和*公平*地共享信道
- 低负载时效率低下：即使只有 1 个活跃节点也分配 1/N 的带宽

随机接入协议
- 低负载时高效：单节点可以充分利用信道
- 高负载时：碰撞开销

轮流协议
- 寻找可以实现上述两种协议优点

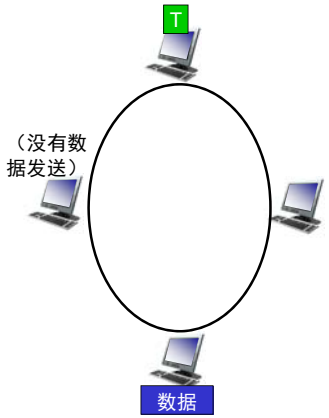
轮流协议

轮询协议：
- 主节点“邀请”其他节点依次进行传输
  - 主节点以循环的方式轮询每个结点。主节点向结点1发送报文告诉它能够传输的帧的最大数量，结点1传输了某些帧后，主节点告诉结点2它可以传输的帧的最大数量. . .
- 缺点：
  - 轮询开销
  - 引入了时延-若只有一个活跃节点，仍然要轮询每个非活跃节点
  - 单点故障（主节点）



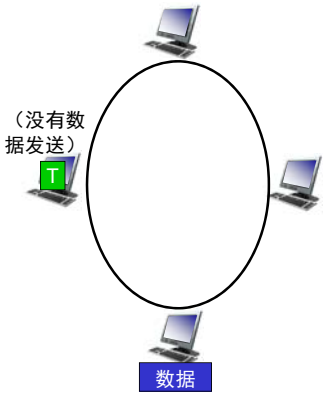
轮流协议

令牌传递协议：
- 该协议没有主节点
- 一个称为*令牌*的小的特殊的帧在节点之间以某种固定的次序进行交换
  - *令牌*按顺序从一个节点传递到下一个节点。
- 当一个节点收到令牌时，仅当它有帧要传输，它发送最大数目的帧数，然后将令牌转发给下一个节点；否则，它立即将令牌转发给下一个节点
- 缺点：
  - 令牌开销
  - 时延
  - 单点故障
    - 一个节点的故障可能崩溃整个信道
    - 节点忘记释放令牌时如何让令牌再次进入循环



轮流协议

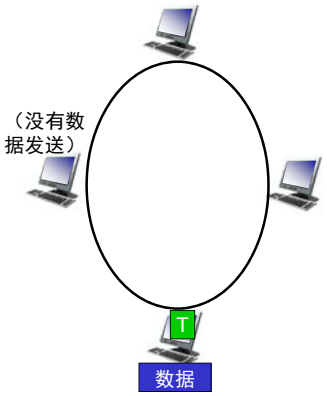
令牌传递协议：
- 该协议没有主节点
- 一个称为*令牌*的小的特殊的帧在节点之间以某种固定的次序进行交换
  - *令牌*按顺序从一个节点传递到下一个节点。
- 当一个节点收到令牌时，仅当它有帧要传输，它发送最大数目的帧数，然后将令牌转发给下一个节点；否则，它立即将令牌转发给下一个节点
- 缺点：
  - 令牌开销
  - 时延
  - 单点故障
    - 一个节点的故障可能崩溃整个信道
    - 节点忘记释放令牌时如何让令牌再次进入循环



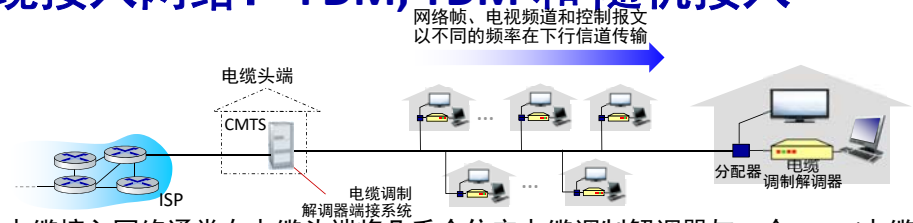
轮流协议

令牌传递协议：

- 该协议没有主节点
- 一个称为 *令牌* 的小的特殊的帧在节点之间以某种固定的次序进行交换
  - *令牌* 按顺序从一个节点传递到下一个节点。
- 当一个节点收到令牌时，仅当它有帧要传输，它发送最大数目的帧数，然后将令牌转发给下一个节点；否则，它立即将令牌转发给下一个节点
- 缺点：
  - 令牌开销
  - 时延
  - 单点故障
    - 一个节点的故障可能崩溃整个信道
    - 节点忘记释放令牌时如何让令牌再次进入循环

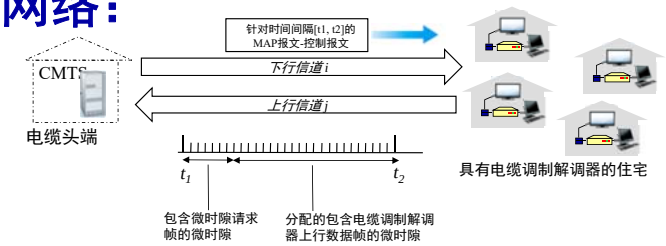


应用多种多路访问协议的例子
电缆接入网络： FDM, TDM 和 随机接入



- 一个电缆接入网络通常在电缆头端将几千个住宅电缆调制解调器与一个CMTS(电缆调制解调器端接系统)连接
- 多个下行（广播）FDM 信道：最大带宽是 1.6 Gbps/信道
  - 仅有一个CMTS（Cable Modem Termination System）向下行信道传输帧，没有碰撞
- 多个上行（广播）信道：最大带宽是 1 Gbps/信道
  - 每条上行信道被划分为时间间隔，每个时间间隔包含一个微时隙序列，CMTS显式地准许各个调制解调器在特定的微时隙中进行帧传输(TMD)
  - 所有电缆调制解调器以随机接入方式竞争使用一部分上行信道时隙，以告知CMTS它要发送数据；若调制解调器在下行控制报文收到对请求分配的响应，则传输帧；否则二进制指数回退；当上行信道上流量很少时，电缆调制解调器可以在分配给它用于请求分配时隙的请求时隙内传输数据帧

电缆接入网络：



DOCSIS: 数据经电缆服务接口规范0-北美标准：定义如何通过电缆调制解调器提供双向数据业务， 主要支持在计算机网与有线电视网之间，以及有线电视前端与用户之间实现IP数据包的传输； 与之相对的是欧洲标准DVB/DAVIC

- FDM 在上行和下行频率信道
- 上行：一些时隙被分配（TMD），一些时隙争用（随机接入）
  - CTMS通过在下行信道上发送 MAP 帧来指定哪个电缆调制解调器能够使用上行微时隙
  - 电缆调制解调器在专用的一组微时隙中向CTMS发送微时隙请求，当推断出有碰撞时使用二进制指数回退将其微时隙请求在以后的时隙中重新发送

MAC协议总结

- 信道划分协议，按时间、频率或编码
  - 时分多址、频分多址、码分多址协议
- 随机接入协议（动态的），
  - ALOHA, 时隙ALOHA, CSMA, CSMA/CD协议
  - 载波侦听：在有线传输中容易做到碰撞检测，但在无线传输中难以做到
  - 以太网使用 CSMA/CD协议
  - 无线网802.11使用 CSMA/CA 协议
- 轮流协议
  - 轮询协议、令牌传递协议
  - 蓝牙传输协议（主设备控制）、 FDDI（光纤分布式数据接口，在光纤上发送数字信号的一组协议，使用双令牌环）、令牌环 IEEE802.5局域网协议（基于令牌传递）

链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

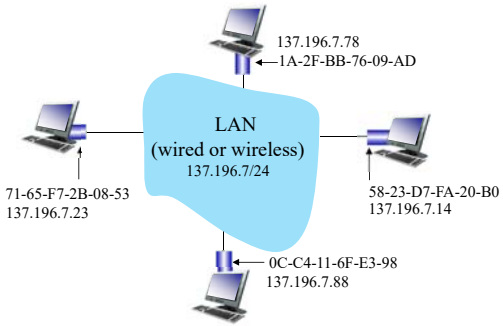
MAC地址

- 32 位 IP 地址:
  - *用于接口的网络层地址*
  - 用于第 3 层（网络层）转发
  - 例如： 128.119.40.136
- MAC地址（也称为局域网地址、物理地址、以太网地址）:
  - 功能: “本地”使用，将“本地”的帧从一个接口连接到另一个物理连接的接口（从IP寻址意义上，本地意思是相同的子网）
  - 大多数局域网使用48 位 MAC 地址（以太网和802.11无线局域网）；被刻入在网络适配器的ROM中，有时也可通过软件设置
  - 例如: 1A-2F-BB-76-09-AD

十六进制记法
（每个"数字"表示4位）

MAC 地址

LAN中的每个接口
- 具有唯一的48位 MAC 地址
- 具有唯一的32 位 本地IP 地址

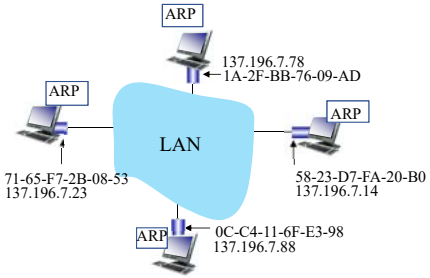


MAC 地址

- MAC 地址分配由 IEEE 管理
- 制造商购买 MAC 地址空间的一部分（以确保唯一性）
  - 购买组成2^{24}个地址的一块地址空间
  - 固定一个MAC地址的前24比特，制造商自己为每个适配器生成后24比特的唯一组合
- 类比:
  - MAC 地址：如社会保险号码
  - IP地址：如邮政地址
- MAC 地址的可移植性
  - 可以将接口从一个局域网移动到另一个局域网
  - IP地址不可移植性：节点的IP地址取决于节点所在的IP子网

ARP：Address Resolution Protocol 地址解析协议

问题：已知接口的IP地址，如何知道接口的MAC地址？



ARP 表: LAN中的每个IP节点 （主机、路由器）都有一个ARP表

- 存储某些LAN节点的IP/MAC地址映射:

  < IP address; MAC address; TTL>

- TTL (Time To Live):寿命值，从表中删除某个映射的时间（典型值为20分钟）

ARP 协议的执行过程

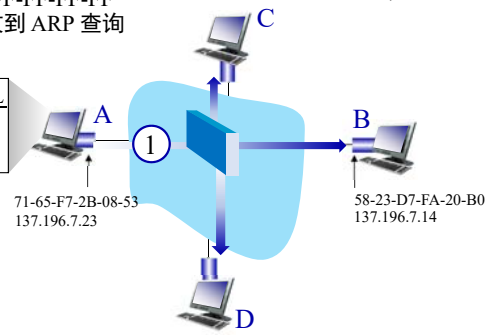
例子：A希望将数据报发送到B

- B的 MAC 地址不在A的 ARP表中，所以A使用ARP协议寻找B的MAC地址

A广播ARP查询，该查询包含 B的IP 地址
1. • 目的 MAC 地址 = FF-FF-FF-FF-FF-FF
   • LAN 中所有的节点都接收到 ARP 查询

A 中的 ARP 表

IP addr	MAC addr	TTL



ARP 协议的执行过程

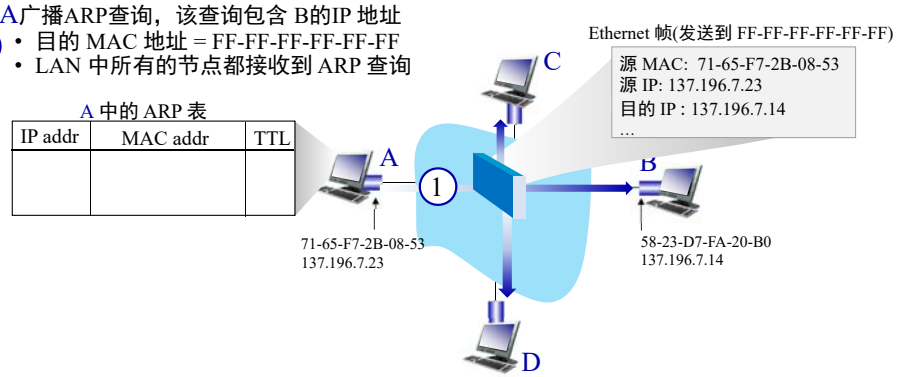
例子：A希望将数据报发送到B

- B的 MAC 地址不在A的 ARP表中，所以A使用ARP协议寻找B的MAC地址

A广播ARP查询，该查询包含 B的IP 地址
1. • 目的 MAC 地址 = FF-FF-FF-FF-FF-FF
   • LAN 中所有的节点都接收到 ARP 查询

A 中的 ARP 表

IP addr	MAC addr	TTL



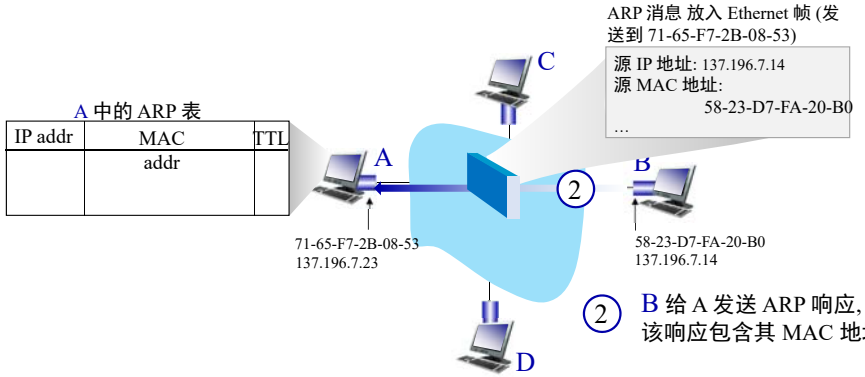
ARP 协议的执行过程

例子：A希望将数据报发送到B

- B的 MAC 地址不在A的 ARP表中，所以A使用ARP协议寻找B的MAC地址

A 中的 ARP 表

IP addr	MAC addr	TTL

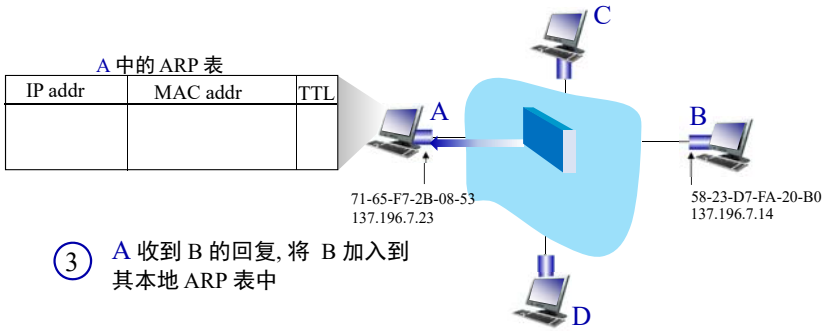


2. B 给 A 发送 ARP 响应，该响应包含其 MAC 地址

ARP 协议的执行过程

例子：A希望将数据报发送到B

- B的 MAC 地址不在A的 ARP表中，所以A使用ARP协议寻找B的MAC地址

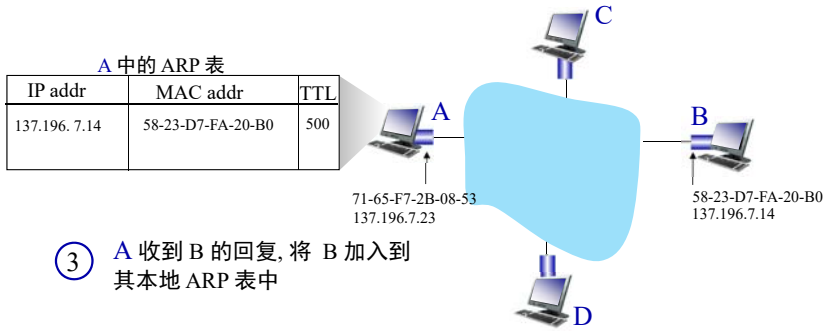


③ A 收到 B 的回复, 将 B 加入到其本地 ARP 表中

ARP 协议的执行过程

例子：A希望将数据报发送到B

- B的 MAC 地址不在A的 ARP表中，所以A使用ARP协议寻找B的MAC地址



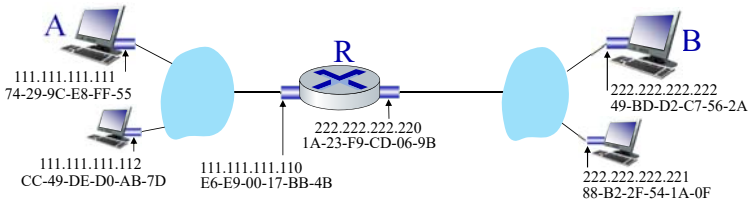
③ A 收到 B 的回复, 将 B 加入到其本地 ARP 表中

发送数据报到子网以外：寻址

示例:通过路由器R将数据报从A发送到B

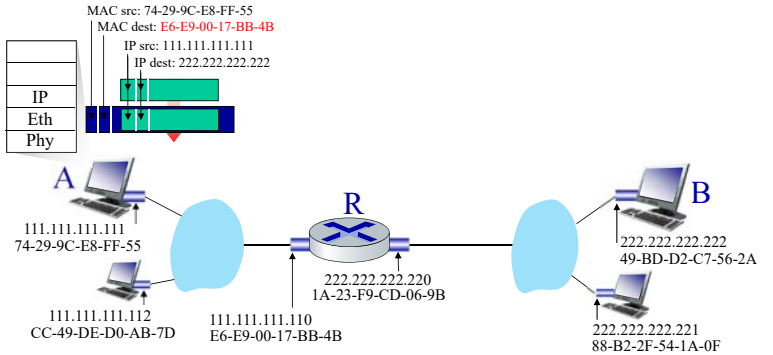
关注在IP（数据报）和MAC层（帧）的寻址

- 假设:
  - A 知道 B 的 IP 地址
  - 知道第一跳路由器R的IP地址(如何知道的?)
  - A 知道 R 的 MAC 地址 (如何知道的?)



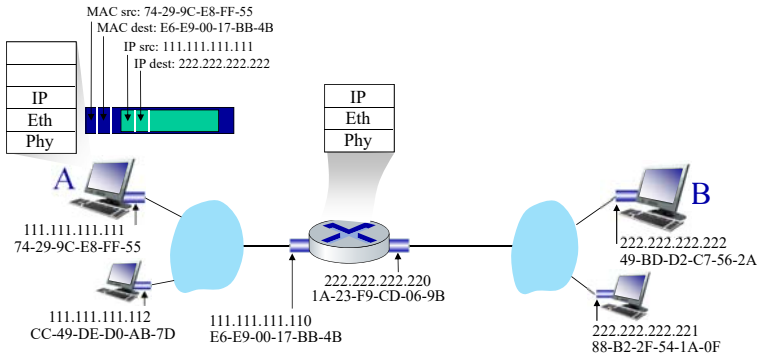
发送数据报到子网以外：寻址

- A 创建 IP 数据报，包含 IP 源 A 和目的 B
- A 创建 链路层帧，包含 A到B 的 IP数据报
  - 问：A向它的网络适配器指示哪个MAC地址作为其目的MAC地址？
  - 答：与A相邻的R 的网络接口的 MAC 地址



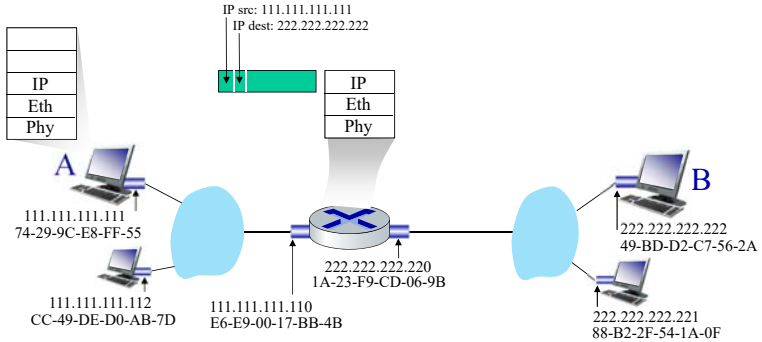
发送数据报到子网以外：寻址

- 帧从 A 发送到 B
- R 接收到帧 , 将数据报取出并传送到它的IP 层



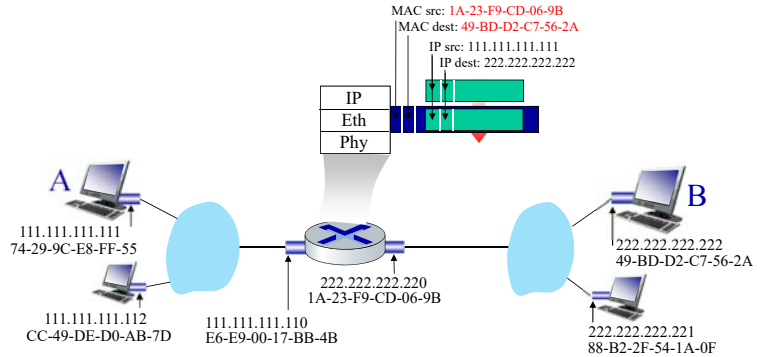
发送数据报到子网以外：寻址

- 帧从 A 发送到 B
- R 接收到帧 , 将数据报取出并传送到它的IP 层



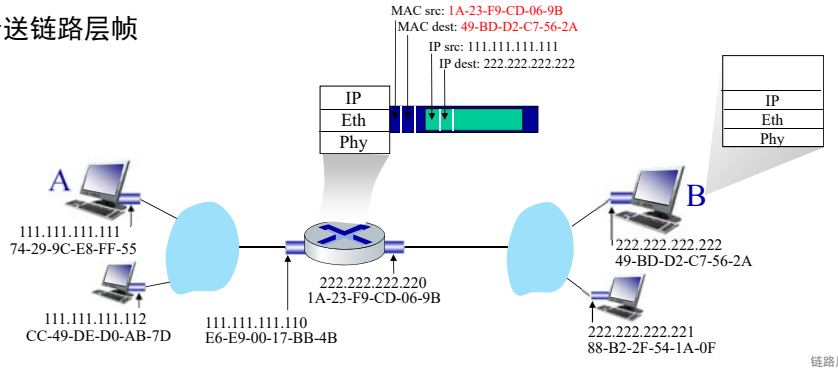
发送数据报到子网以外：寻址

- R 确定传出接口，将IP源为A、目的为B的数据报发送到链路层
- R 创建链路层帧，包含 A 到 B 的 IP数据报，帧的目的地址为 B的 MAC地址



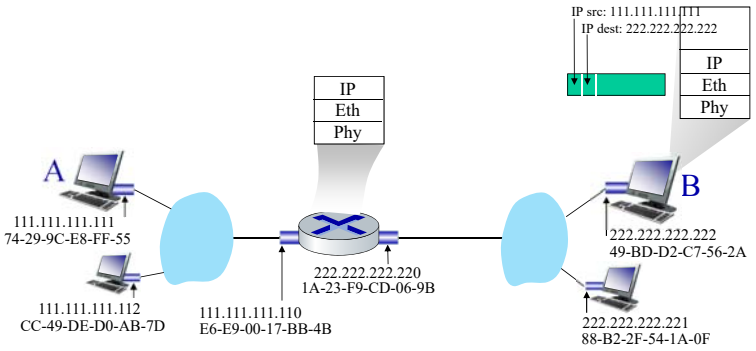
发送数据报到子网以外：寻址

- R 确定传出接口，将IP源为A，目的为B的数据报发送到链路层
- R 创建链路层帧，包含 A 到 B 的 IP数据报，帧的目的地址为 B 的 MAC地址
- 传送链路层帧



发送数据报到子网以外：寻址

- B 接收到帧，提取 IP数据报的目的地址 B
- B 将数据报发送到 IP 层



链路层和局域网: 路线图

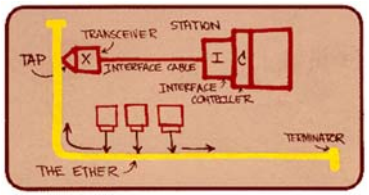
- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- **局域网**
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

以太网 Ethernet

最著名的有线LAN技术:

- 广泛部署的局域网技术
- 更简单，便宜
- 速率: 10 Mbps – 400 Gbps
- 单芯片，多速率(如：博通（Broadcom）公司的 BCM5761将三速 IEEE 802.3（ 10/100/1000BASE-T） 兼容的MAC、PCI Express总线接口、缓存、物理层收发器组合在单个芯片上

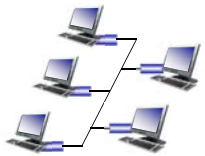


梅特卡夫以太网框架（1973年在施乐实习，1975年梅特卡夫和博格斯发表了论文“以太网：本地计算机网络的分布式分组交换”）

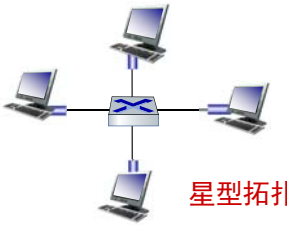
https://www.uspto.gov/learning-and-resources/journeys-innovation/audio-stories/defying-doubters

以太网的物理拓扑

- 总线拓扑: 从20世纪80年代到90年代中期一直流行
  - 所有节点在同一个冲突域中（可以发生冲突）
- 星型拓扑: 从20世纪90年代后期至今流行
  - 目前的部署：中心是二层交换机（以前中心是物理层的集线器）
  - 交换机将链路彼此隔离，局域网中不同链路可以以不同速率在不同媒体上运行。



总线拓扑：同轴电缆



星型拓扑：交换机

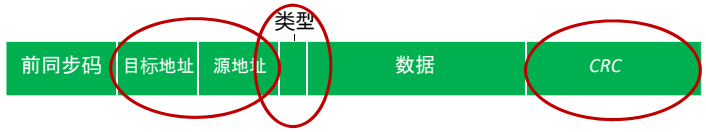
以太网帧结构

发送接口将IP 数据报（或其他网络层协议的数据包）封装为以太网帧

前同步码:

- 用于同步接收方和发送方的时钟速率。
- 前七个字节为10101010，最后一个字节为10101011

以太网帧结构



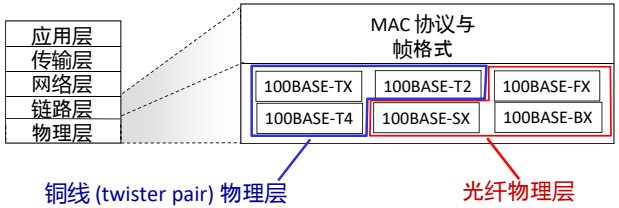
- 地址：源网络适配器或目标网络适配器的6字节MAC地址
  - 适配器收到一个帧时，若该帧的目的地址是适配器的MAC地址或是广播MAC地址，那么适配器将把该帧传递给本机的网络层协议。
  - 否则，适配器丢弃该帧。
- 类型：标识高层协议
  - 多数情况下是IP协议，但是也有可能是其他协议，如Novell IPX, AppleTalk
  - 用于复用多种网络层协议。
- CRC：在接收方进行循环冗余检测
  - 若检测到错误，则丢弃该帧

以太网：不可靠，无连接

- 无连接: 发送网卡和接收网卡之间无需握手
- 不可靠: 接收网卡不向发送网卡发送ACK或NAK
  - 只有当原始发送者使用了更高层的协议（如TCP)时，丢弃帧中的数据才会被重发，否则丢弃的数据就丢掉了。
- 以太网的MAC协议: 无时隙的、使用二进制指数回退的CSMA/CD

802.3（有线）以太网标准：链路层 & 物理层

- 802. 3标准制定了以太网的技术标准，包括物理层连线、电子信号和媒体访问协议等内容
- 许多不同的以太网标准
  - 使用相同的MAC协议和帧格式
  - 不同的速度: 2 Mbps, 10 Mbps, 100 Mbps, 1Gbps, 10 Gbps, 40 Gbps
  - 不同的物理层介质：光纤，铜线



链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

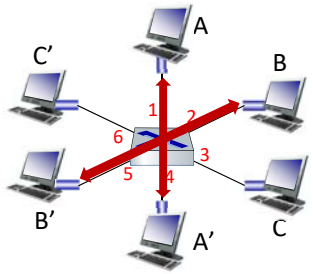
- Web页面请求的历程

以太网交换机

- 交换机是一种链路层设备(通常情况下): 扮演一个主动的角色
  - 存储/转发以太网帧
  - 检查传入帧的MAC地址, 选择性地将帧转发到一个或多个出链路中，使用 CSMA/CD 访问网段
- 透明的: 主机并不知道交换机的存在
- 即插即用, 自学习
  - 交换机无需配置

交换机: 多个同时传输

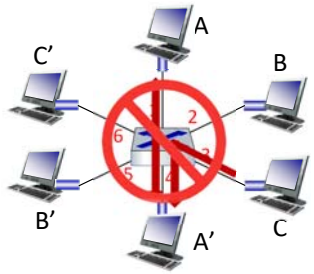
- 主机与交换机之间有专用的、直接的连接。
- 交换机缓存数据包
- 每个入链路上都使用了以太网协议，因此:
  - 无碰撞；全双工
  - 每条链路是它自己的冲突域
- 交换: A-to-A′ 和B-to-B′ 可以同时传输，不会发生碰撞



有6个接口(1,2,3,4,5,6) 的交换机

交换机: 多个同时传输

- 主机与交换机之间有专用的、直接的连接。
- 交换机缓存数据包
- 每个入链路上都使用了以太网协议，因此:
  - 无碰撞；全双工
  - 每条链路是它自己的冲突域。
- 交换: A-to-A′ 和B-to-B′ 可以同时传输，不会发生碰撞。
  - 但是A-to-A′ 和C to A′ 不能同时传输



有6个接口(1,2,3,4,5,6)的交换机

交换机表

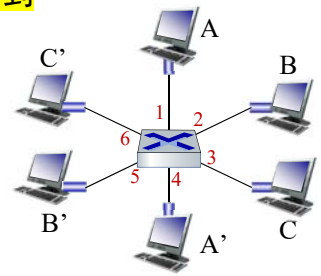
Q: 交换机是如何知道，A'可以通过接口4到达，而B'可以通过接口5到达呢？

A: 每个交换机都有一个交换机表, 每个表项:

- 包含①一个MAC地址, ②通向该MAC地址的交换机接口, ③表项放置在表中的时间
- 与路由表类似!

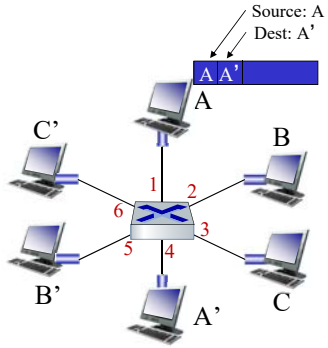
Q: 在交换机表中，表项是如何创建和维护的?

- 是否类似于路由协议?



交换机: 自学习

- 交换机学习 "哪个接口可以到达哪个主机"，即主机与接口的对应关系。
  - 收到帧时，交换机将学习发送者的"位置"，即其所在的局域网网段
  - 在交换机表中记录发送方MAC地址及对应的位置（接口）信息。

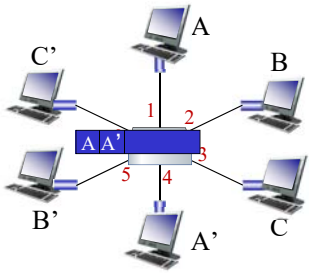


MAC addr	interface	TTL

交换机表
(初始为空)

交换机: 自学习

- 交换机学习 "哪个接口可以到达哪个主机"，即主机与接口的对应关系。
  - 收到帧时，交换机将学习发送者的"位置"，即其所在的局域网网段
  - 在交换机表中记录发送方MAC地址及对应的位置（接口）信息。



MAC addr	interface	TTL
A	*1*	*60*

交换机表
(初始为空)

交换机: 帧过滤与转发

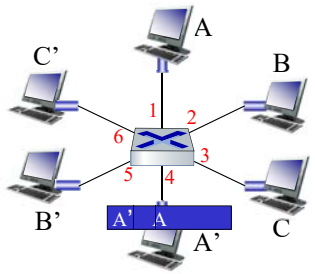
过滤：决定一个帧应该转发到某个接口还是应当将其丢弃
转发：决定一个帧应该被导向哪个接口

当帧达到交换机时:

1. 记录到达的接口x以及发送主机的MAC地址
2. 使用目的MAC地址对交换机表进行索引
3. if 为目的地址找到对应表项
   then {
   if 目的地址对应的接口为x
   then 丢弃该帧
   else 将该帧转发到表项指定的接口
   }
   else 洪泛 /* 即将该帧转发到除x外的所有接口*/

自学习, 转发: 例子

- 目的主机A' 的位置未知:洪泛
- 目的主机A的位置已知：选择相应的链路，进行发送

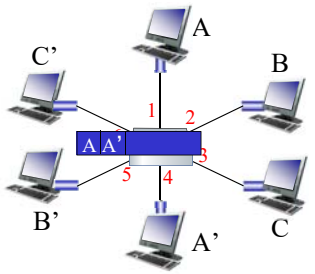


MAC addr	interface	TTL
A	*1*	*60*

交换机表
(初始为空)

自学习, 转发: 例子

- 目的主机A' 的位置未知:洪泛
- 目的主机A的位置已知：选择相应的链路，进行发送

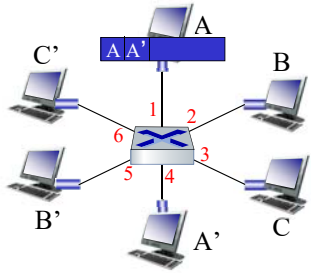


MAC addr	interface	TTL
A	*1*	*60*
A'	*4*	*60*

交换机表
(初始为空)

自学习, 转发: 例子

- 目的主机A' 的位置未知:洪泛
- 目的主机A的位置已知：选择相应的链路，进行发送

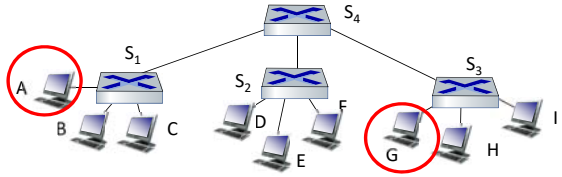


MAC addr	interface	TTL
A	*1*	*60*
A'	*4*	*60*

交换机表
(初始为空)

互连的交换机

自学习的交换机可以连接在一起:

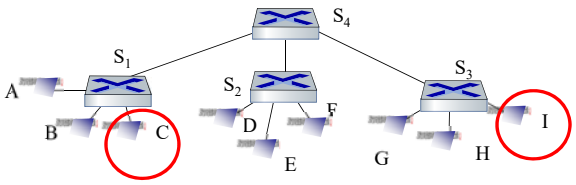


Q: 主机A向主机G发送数据时，S_1是如何知道帧要通过S_4和S_3转发的呢？

- *A:* 自学习！（工作原理与单交换机的情况完全相同!）

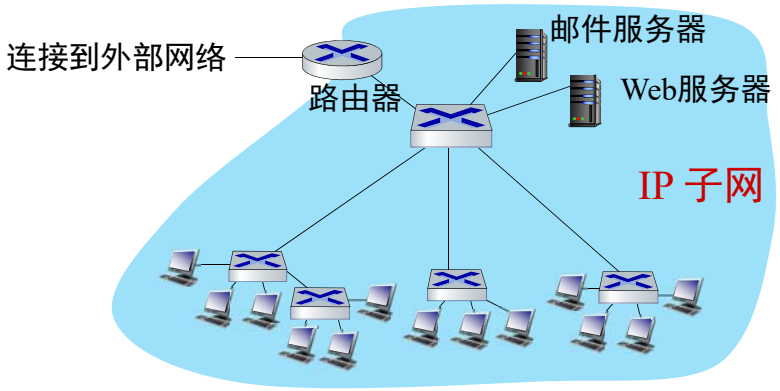
多交换机自学习的例子

假定主机C向I发送帧, I 向C发送响应



Q: 请给出S_1, S_2 , S_3, S_4中的交换机表和帧转发情况

常见的小型机构的网络拓扑结构



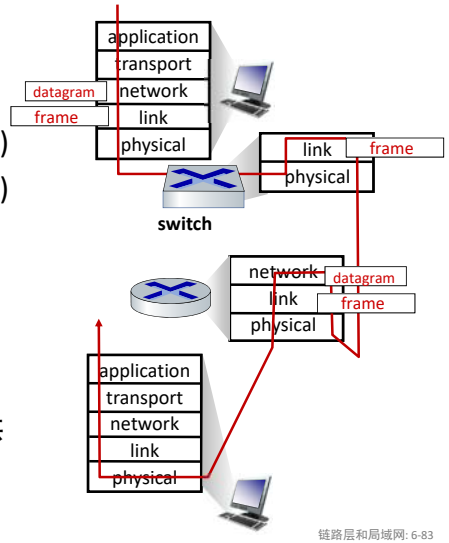
交换机和路由器比较

都是存储转发设备:

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

都有转发表:

- 路由器: 使用网络层地址转发分组的存储转发，使用路由算法计算路由表
- 交换机: 使用MAC地址转发帧，使用洪泛（flooding）、自学习交换机表



链路层和局域网: 路线图

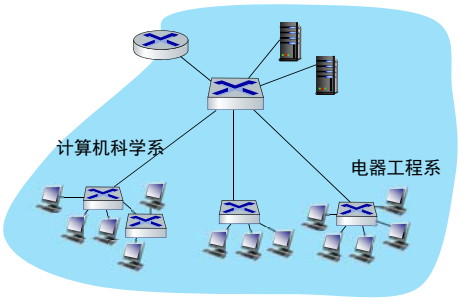
- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

虚拟局域网(VLAN): 动机

现代机构的局域网通常配置为等级结构，每个工作组或部门有自己的交换局域网，经过一个交换机等级结构与其他工作组的交换局域网互联

Q: 当局域网规模扩大时，或者一个用户在组织内部的不同工作组之间移动，会发生什么？

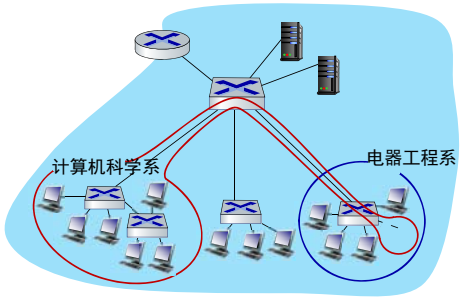


一个广播域：

- *缺乏流量隔离：* 所有2层广播流量(携带ARP、DHCP报文或那些目的地还没有被自学习交换机学习到的帧) 必须跨越整个局域网
- *交换机的无效使用：* 无法隔离不同组的流量

虚拟局域网(VLAN): 动机

Q: 当局域网规模扩大时，或者一个用户在组织内部的不同工作组之间移动，会发生什么？



一个广播域：

- *缺乏流量隔离：* 所有2层广播流量(携带ARP、DHCP报文或那些目的地还没有被自学习交换机学习到的帧) 必须跨越整个局域网
- *交换机的无效使用：* 无法隔离不同组的流量

管理问题：

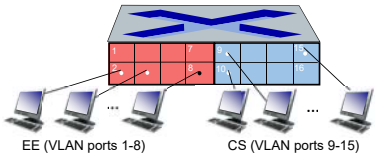
- 计算机科学系的老师办公室搬到电器工程系 *物理上* 连接点为EE的交换机，但是想在逻辑上保持CS交换机的连接
- *难以管理用户：用户*在不同组之间移动、或者一用户属于多个组

基于端口的虚拟局域网

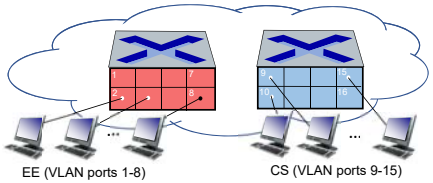
虚拟局域网(VLAN)

支持VLAN的交换机可以在单个物理局域网的基础设施上配置多个虚拟局域网。

基于端口的 VLAN: 通过交换机管理软件对交换机端口进行分组，使得单个物理交换机 ……

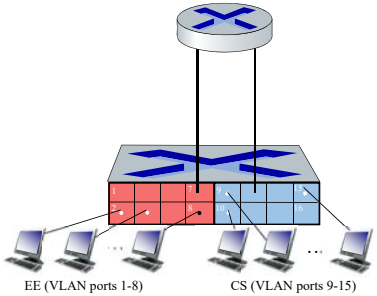


作为多个虚拟交换机运行。

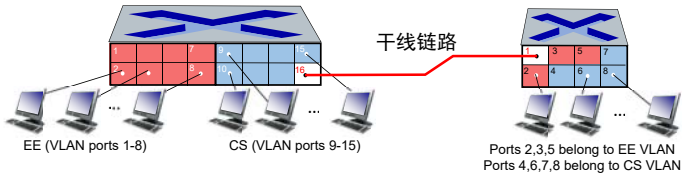


基于端口的虚拟局域网

- 流量隔离：每个VLAN的端口形成一个广播域，即来自一个端口的广播流量仅能达到该组中的其他端口；不同VLAN的帧彼此隔离
  - 也可以定义基于MAC地址而不是交换机端口的VLAN
- 动态成员配置：端口可以在VLAN间动态配置
- VLAN间的转发：彼此隔离的VLAN如何通信呢？- 路由转发
- 现实应用中，厂商销售的是带有VLAN交换机和路由器的单一设备



跨多交换机的VLAN

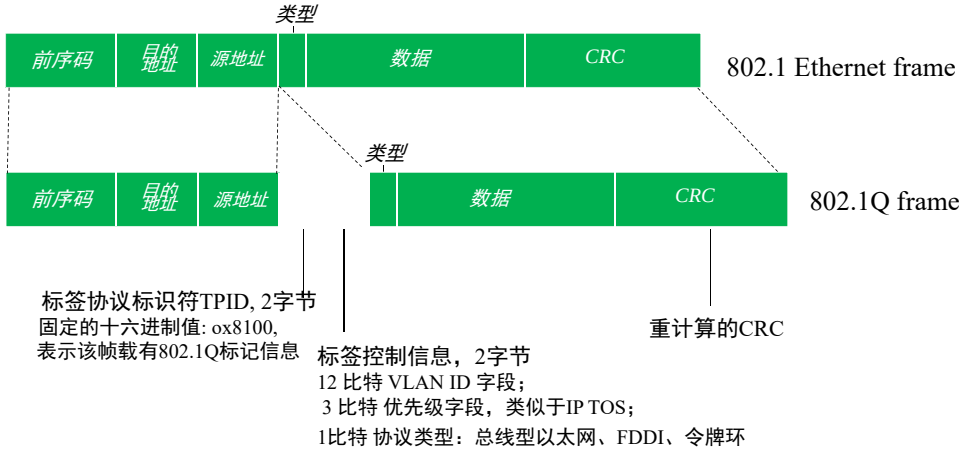


VLAN干线连接(VLAN Trunking)：每台交换机上有一个特殊端口被配置为干线端口（Trunk Port）

trunk port: 在跨多个交换机上的VLAN间传输帧

- 交换机之间转发的同一个VLAN的帧不是普通的802.1帧 (必须携带VLAN ID信息)
- 802.1q 协议定义了扩展的以太网帧格式，为trunk port转发的这些帧增加/移除额外的首部字段

802.1Q VLAN 帧格式



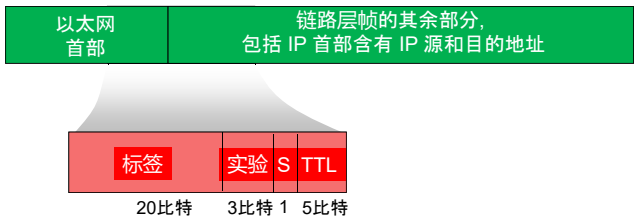
链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- LANs
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化: MPLS
- 数据中心网络

- Web页面请求的历程

多协议标签交换
Multi-Protocol Label Switching (MPLS)

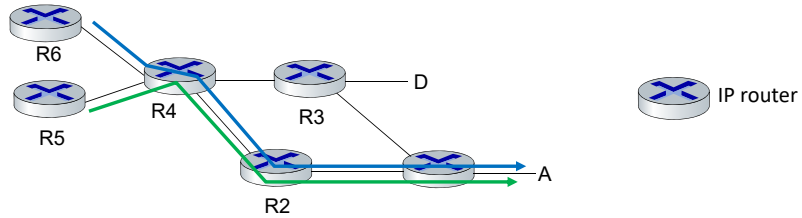
- 目标: 进行高速IP数据报转发，使用固定长度标签 (而不是目的地IP地址)转发数据报
  - 使用固定长度标识符更快的查找
  - 借鉴了虚电路方法
  - IP数据报仍然保留IP地址!
  - 一个MPLS加强的帧仅能在两个均为MPLS使能的路由器间发送



MPLS使能的路由器

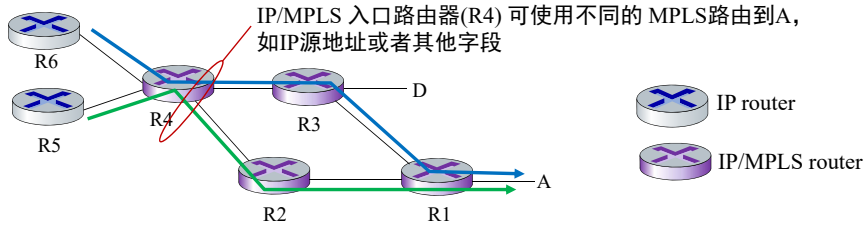
- 又称标签交换路由器
- 只根据标签值转发报文到出口(*不检查IP地址*)
  - MPLS转发表与IP转发表不同
- *灵活性:* MPLS转发决策可以与基于IP的转发决策不同
  - 实现流量工程：如对于到相同目的地址的网络流，由于源地址的不同而路由不同
  - 链路出现问题时进行快速重新路由：提前计算的备份路径

基于MPLS标签 与 基于IP地址的转发



- 基于IP地址转发: 到目的地的路径仅由目的地地址决定

基于MPLS标签 与 基于IP地址的转发



- 基于IP地址转发: 到目的地的路径仅由目的地地址决定
- 基于MPLS标签转发: 到目标的路径可以基于源地址和目标地址
  - 通用转发(10年前的MPLS)
  - *快速重路由：*执行MPLS转发路径的快速恢复 - 提前计算备份路由，以防链路故障

MPLS使能的路由器

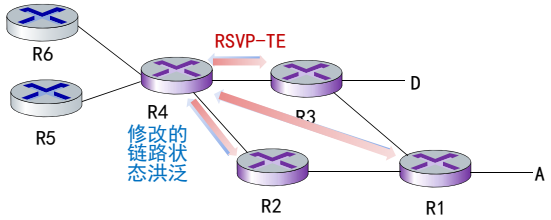
- 又称标签交换路由器
- 只根据标签值转发报文到出口(*不检查IP地址*)
  - MPLS转发表与IP转发表不同
- *灵活性:* MPLS转发决策可以与基于IP的转发决策不同
  - 实现流量工程：如对于到相同目的地址的网络流，由于源地址的不同而路由不同
  - 链路出现问题时进行快速重新路由：提前计算的备份路径

Q1：路由器如何知道它的邻居是否是MPLS使能的呢？

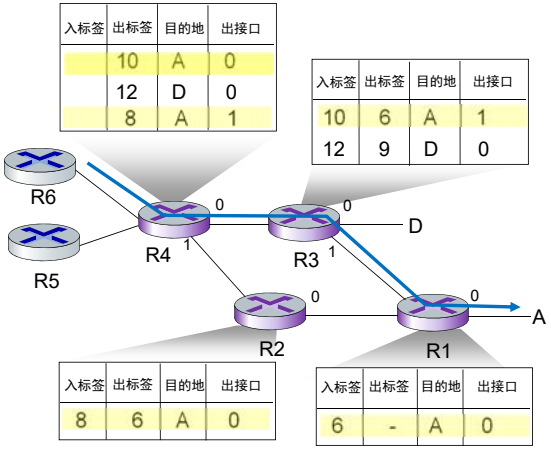
Q2：路由器如何知道哪个标签与给定IP目的地相联系呢？

支持MPLS的路由选择协议

- 扩展OSPF、IS-IS链路状态路由选择协议，使之承载MPLS路由选择信息：
  - 如，链路带宽、保留的链路带宽的数量
- 起始的MPLS使能路由器使用RSVP-TE协议在下游MPLS使能路由器上建立基于MPLS标签的转发



MPLS 增强的转发



链路层和局域网: 路线图

- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- LANs
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化：MPLS
- 数据中心网络

- web请求生命中的一天

数据中心网络

构建了大量的数据中心：成千上万的主机，在很近的距离紧密耦合的：

- 电子商务 (如， Amazon)
- 内容服务商 (如，YouTube, Akamai, Apple, Microsoft)
- 搜索引擎，数据挖掘(如，Google)

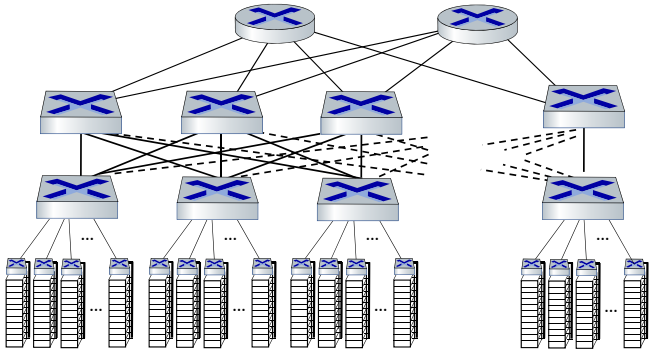
每个数据中心都有自己的数据中心网络

挑战：

- 多个应用程序，每个应用程序为大量客户服务
- 如何保障可靠性
- 如何进行管理与负载均衡，避免处理、网络和数据瓶颈

在一个40英尺高的集装箱里的微软芝加哥数据中心

数据中心网络



边界路由器
- 将数据中心网络与互联网相连

第一层交换机
- 连接~16 第二层交换机

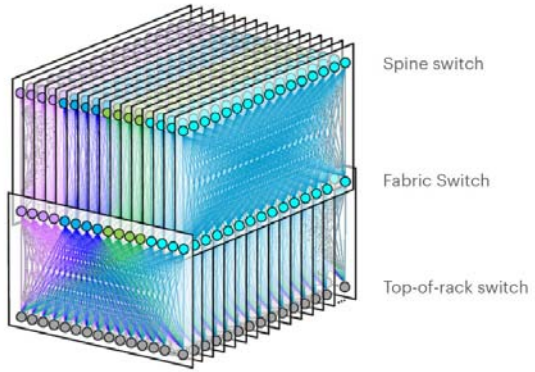
第二层交换机
- 连接~16个机架顶部交换机

机架顶部（TOR, Top of Rack）交换机
- 每一个机架顶部有一台交换机
- 与每台刀片有40-100Gbps 以太网连接

服务器机架
- 数据中心的主机称为刀片
- 每机架堆放20- 40台刀片

数据中心网络的例子

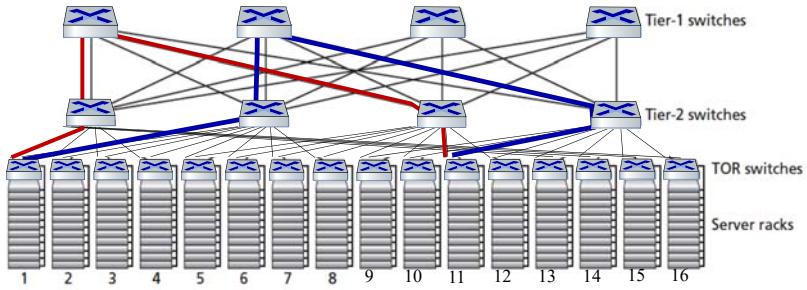
Facebook数据中心网络拓扑结构，称为F16：



https://engineering.fb.com/data-center-engineering/f16-minipack/ (posted 3/2019)

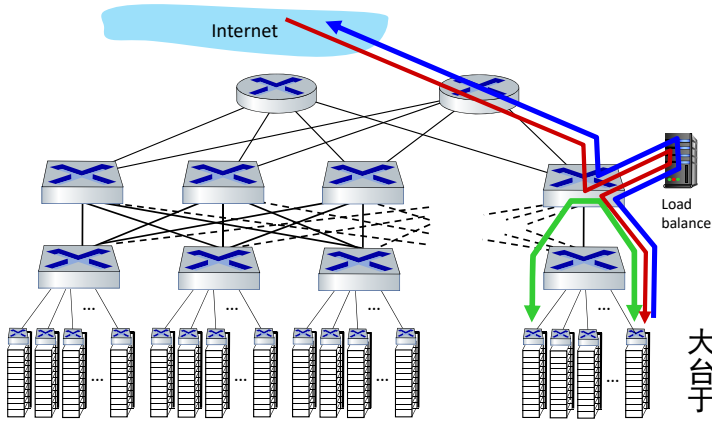
数据中心网络：多路径网络连接

- 交换机和机架之间存在非常丰富的连接:
  - 增加机架之间的吞吐量(因为存在多条路由路径)
  - 通过冗余提高了可靠性



图上标红和蓝的两条线的为机架1和机架11之间不相交的两条路径

数据中心网络： 应用层路由



负载均衡：应用层路由
- 接收外部客户端的请求
- 将请求分发给数据中心内某一台主机上（该主机可能再调用其他主机的服务来处理该请求）
- 返回数据给外部客户端(对客户端隐藏数据中心内部)

大型的数据中心通常有几台负载均衡器，每台服务于特定的云服务

数据中心网络: 协议创新

- 链路层:
  - RoCE（RDMA over Converged Ethernet）: 基于融合以太网的远端内存直接访问：数据在网络中两个节点的应用程序的虚拟内存间传输、不需要额外的复制和缓存传输、不需要内核参与、传输的所有处理都由NIC硬件中完成。
- 传输层:
  - 传输层拥塞控制使用ECN(explicit congestion notification) (DCTCP, DCQCN)
  - 使用逐跳的拥塞控制实验方法
- 路由选择、管理:
  - SDN被广泛使用于各个组织内和组织间的数据中心
  - 尽可能将相关服务，数据放置在尽可能近的位置（例如，在同一机架或附近机架中），以最大程度地减少第二层和第一层的通信

链路层和局域网: 路线图

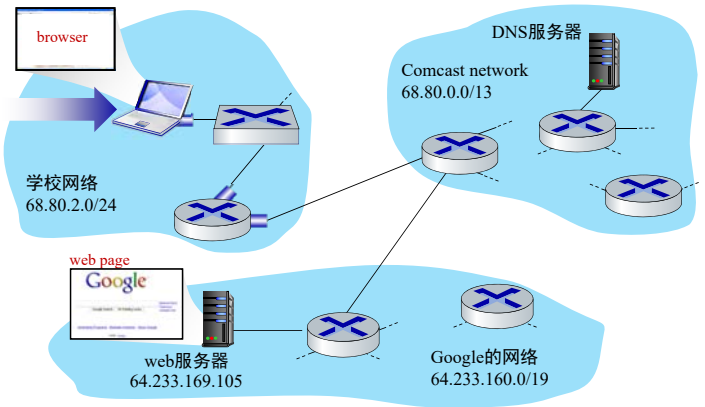
- 链路层概述
- 差错检测和纠正技术
- 多路访问链路和协议
- LANs
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs
- 链路虚拟化：MPLS
- 数据中心网络

- Web页面请求的历程

综合：Web页面请求的历程

- 我们沿网络协议栈向下的旅程完成啦!
  - 应用层、传输层、网络层、链路层
- 将学过的内容放在一起!
  - 目标：在一个看似简单的场景中、理解其中所涉及的所有层的协议
  - 场景：某学生将笔记本电脑连接到校园网络，请求www.google.com网页

Web页面请求的历程: 场景

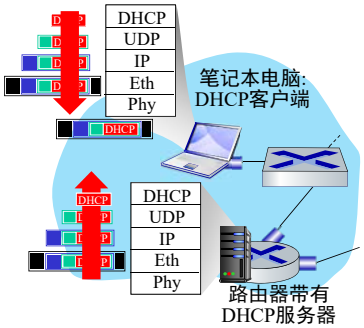


场景:

- 笔记本连接到网络
- 请求web页面 www.google.com

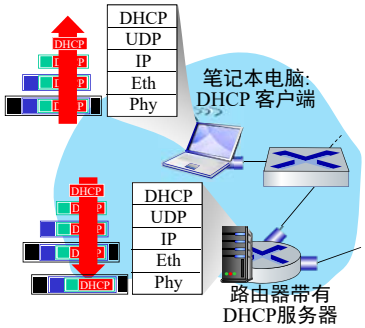
听起来很简单!

Web页面请求的历程: 连接到网络



- 连接到网络的笔记本需要获取自己的IP地址，第一跳路由的IP地址，DNS服务器的地址：使用?协议 DHCP
- DHCP请求报文从上至下会被封装为UDP报文，IP报文，802.3以太网报文
- 以太网帧在LAN上广播（目的地址:FFFFFFFFFFFF)，被路由器上的DHCP服务器接收到
- 以太网报文解封装为IP报文，UDP报文，DHCP报文

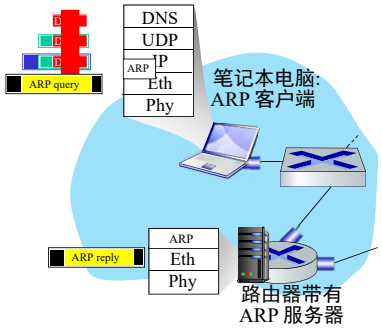
Web页面请求的历程: 连接到网络



- DHCP服务器构造DHCP ACK报文，里面包含客户端的IP地址，第一跳路由的IP地址，DNS服务器的名字和IP地址
- 报文在DHCP服务器封装，通过LAN转发，客户端接收到后解封装
- DHCP客户端接收到DHCP ACK回应报文

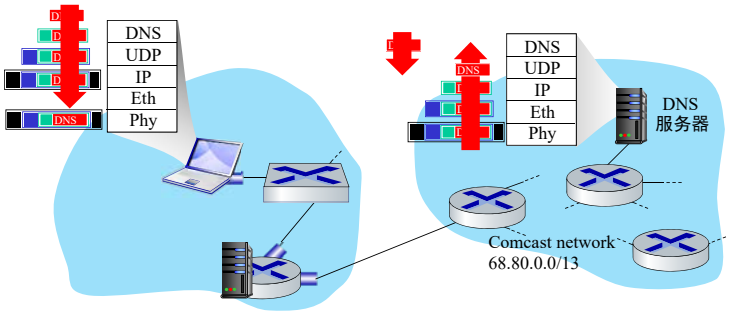
现在客户端拥有IP地址，DNS服务器的名字和地址，第一跳路由的IP地址

Web页面请求的历程: ARP (before DNS, before HTTP)



- 在发送HTTP请求之前，需要www.google.com的IP地址: DNS
- DNS请求被依次封装到UDP、IP、Eth报文中。为了发送给路由器，需要知道路由器的MAC地址 : ARP
- 发送ARP 请求广播，被路由器接收到，返回带有路由器MAC地址的ARP响应报文
- 现在客户端知道第一跳路由的MAC地址，可以发送包含DNS请求的以太帧

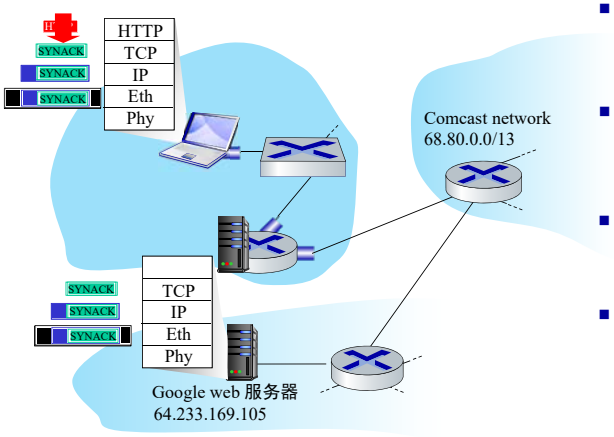
Web页面请求的历程: 使用DNS



- IP数据报文解封装到DNS报文
- DNS服务器将www.google.com的IP地址回复给客户端

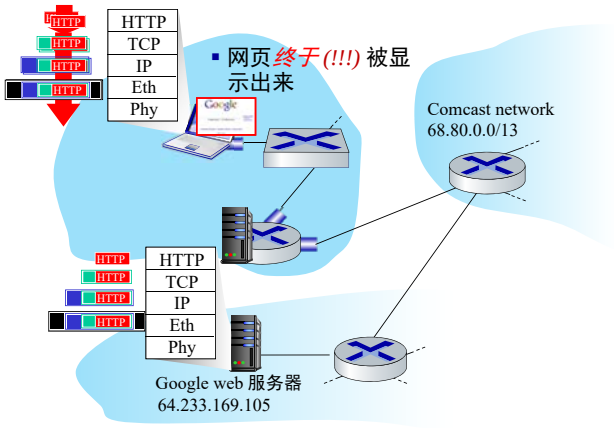
- 包含DNS请求的IP数据报文通过LAN交换机从客户端转发到第一跳路由
- IP数据报文从校园网络转发到运营商(Comcast)网络，被路由到DNS服务器（由RIP，OSPF，IS-IS和/或BGP路由协议创建的路由表转发)

Web页面请求的历程：TCP 和HTTP



- 为了发送HTTP请求，客户端首先创建到Web 服务器的TCP socket
- TCP SYN报文段(segment) (tcp三次握手的第一步) 被路由到web服务器
- Web 服务器回复 TCP SYN/ACK (tcp三次握手的第二步)
- 包含TCP SYN/ACK的报文段最终达到笔记本的以太网网卡；数据报在操作系统中分解到相应的TCP套接字，从而进入连接状态！

Web页面请求的历程: HTTP 请求和响应



- HTTP 请求通过TCP socket 发送
- 包含HTTP请求的IP 数据报文 被路由到 www.google.com
- web 服务器回复HTTP响应报文 (包含web页面)
- 包含HTTP响应报文的IP 数据报文被路由回客户端

第六章：链路层和局域网总结

- 链路层服务背后的原理:
  - 差错检测和纠正技术
  - 共享广播信道: 多路访问
  - 链路层寻址
- 各种链路层技术的实例化与实现
  - 以太网
  - 交换局域网（中心是交换机的星型拓扑）、虚拟局域网（VLAN）
  - 虚拟化网络作为链路层：多协议标签交换（ MPLS ）网络 – 以提高流量管理能力、转发路径的快速恢复
  - 数据中心网络
- 综合: Web页面请求的历程

链路层和局域网: 完成！

- 链路层概述
  - 链路层提供的服务
  - 链路层在何处实现
- 差错检测和纠正技术
- 多路访问链路和协议
- 局域网
  - 寻址, ARP
  - Ethernet
  - 交换机
  - VLANs

- 链路虚拟化: MPLS
- 数据中心网络
- Web页面请求的历程

第六章 作业

- 第8版
- 2、5、11、14、15、17、21、26、29、32

- 第7版
- 2、5、11、14、15、17、21、26、29、32