链路层和局域网

链路层概述

一些术语

1. 节点(node):运行链路层协议的任何设备，包括主机、路由器、交换机和WiFi接入点
2. 链路(link):沿着通信路径连接相邻节点的通信信道
3. 第二层协议以数据单元帧(frame)，封装数据报
4. 数据链路层负责从一个节点通过链路将(帧中的)数据报发送到相邻的物理节点
5. 数据报(分组)在不同的链路上以不同的链路协议传送：
    1. 第一跳链路：以太网
    2. 中间链路：帧中继链路
    3. 最后一条802.11
    4. 不同链路协议提供不同的服务

链路层提供的服务

1. 成帧(framing):在每个网络层数据报经链路传送之前，链路层协议都将其用链路层帧封装起来。一个帧由一个数据字段和若干首部字段组成
2. 链路接入：媒体访问控制(Medium Access Control,MAC)协议规定了帧在链路上传输的规则。
3. 可靠交付：保证无差错地经链路层移动每个网络层数据报，通过确认和重传实现。
4. 差错检测和纠正：前者通过让发送节点在帧中包括差错检测比特，让接收节点进行差错检查。后者区别在于接收方不仅能检测帧中出现的比特差错，而且能够准确地确定帧中地差错出现的位置。
5. 流量控制：使得相邻的发送和接收方节点的速度匹配
6. 差错纠正：接收端检查和纠正bit错误，不通过重传纠正错误

链路层在哪里实现

1. 在每一个主机上，每个路由器上和交换机的每一个端口上
2. 链路层功能在网络适配器(network adapter,又称网络接口卡，Network Interface Card，NIC)实现或者在一个芯片组上
3. 接到主机的系统总线上
4. 硬件、软件和固件的综合体

适配器通信

1. 发送方：
    1. 在帧中封装数据报
    2. 加上差错控制编码，实现RDT和流量控制功能等
2. 接收方
    1. 检查有无出错，执行RDT和流量控制功能等

    2. 解封装数据报，并将其交付到上层

差错检测和纠正技术

比特级差错检测和纠正(bit-level error detection and correction)：从一个节点发送到另一个物理上连接的邻近节点的链路层帧中的比特损伤进行检测和纠正

差错检测和纠正比特(Error-Detection and Correction,EDC):差错检测和纠正位(冗余位) 数据D:数据由差错检测保护，可以包含头部字段

未检出比特差错(undetected bit error):接收方可能无法知道接收的信息中包含比特差错。

前向纠错(Forward Error Correction,FEC)：接收方检测和纠正差错的能力

三种检测差错的技术

1. 奇偶校验
    1. 单bit奇偶校验(parity bit)：检测单个bit级错误/奇数个比特差错
    2. 二维奇偶校验(two-dimensional parity)：检测和纠正单个bit错误
2. 检验和方法
    1. 目标：检测在传输报文段时的错误(如位翻转)
    2. 发送方：
        1. 将报文段看成16-bit整数
        2. 报文段的校验和：和的反码形成了携带在报文段首部的Internet检验和(Internet checksum)
        3. 计算接收到的报文段的校验和
        4. 检查是否与携带校验和字段值一致：
            1. 不一致，检测错误
            2. 一致，还是可能有错误
        5. 更简单的方法：检测结果是否为全1比特来检测检验和
3. 循环冗余检测(Cyclic Redundancy Check,CRC)
    1. CRC编码又称多项式编码(polynomial code),是强大的差错检测码
    2. 将数据比特D，看成是二进制的数据
    3. 生成多项式G：双方协商r+1位模式(r次方)
        1. 生成和检查所使用的位模式
    4. 目标：选择r位CRC附加位R，使得
        1. d+r比特正好被G整除，即余数为0
        2. 若非0则检查错误
        3. R=remainder[D2^r/G]

多路访问链路和协议

两种类型的链路(一个子网内部链路连接形式)

1. 点对点链路(point-to-point link):有链路一端的单个发送方和链路另一端的单个接收方组成；exam.点对点协议(point-to-point protocol,PPP),高级数据链路控制(high-level data link control,HDLC)
2. 广播链路(broadcast link)：能让多个发送和接收节点都连接到相同的、单一的、共享的广播信道上。当任何一个节点传输一个帧时，信道广播该帧，每个其他节点都受到一个副本。exam.传统以太网、HFC上行链路、802.11无线局域网

多路访问协议(multiple access protocol)：节点通过这些协议来规范它们在共享的广播信道上的传输行为。

1. 单个共享的广播型链路
2. 所有节点同时多个帧，即传输的帧在所有的接收方处碰撞(collide)
3. 三种多路访问协议：
    1. 信道划分协议(channel partitioning protocol)
        1. 把信道划分成小片(时间、频率、编码)
        2. 分配片给每个节点专用
    2. 随机接入协议(random access protocol)
        1. 信道不划分，允许碰撞(反复重发)
        2. 冲突后恢复
    3. 轮流协议(taking-turns protocol)
        1. 节点依次轮流
        2. 很多数据传输的节点可以获得较长的信道使用权

信道划分MAC协议：TDMA(time division multiple access)

1. 轮流使用信道，信道的时间分为周期
2. 每个站点使用每周期中固定的时隙(长度=帧传输时间)传输帧
3. 站点无帧传输，时隙空闲->造成浪费

信道划分MAC协议：FDMA(frequency division multiple access)

1. 信道的有效频率范围被分成一个个小的频段
2. 每个站点被分配一个固定的频段
3. 分配给站点的频段如果没有被使用，则空闲

编码多路访问CDMA(Coe Division Multiple Access,CDMA)

1. 所有站点在整个频段上同时进行传输，采用编码原理加以区分
2. 完全没冲突
3. 对面每个节点分配一种不同的编码，每个节点用它唯一的编码对它发送的数据进行编码

随机存取协议

1. 当节点有帧要发送时
    1. 以信道带宽的全部Rbps发送
    2. 没有节点间的预先协调
2. 两个或更多节点同时传输，会发生碰撞
    1. 有碰撞时，设计碰撞的每个节点反复地重发它的帧(分组)，直到该帧无碰撞的通过为止
    2. 不立即重发，而是在重发之前等待一个随机时延，时延是独立选择的

时隙ALOHA协议

1. 假设
    1. 所有帧是登场的
    2. 时间被划分为相等的时隙，每个时隙可发送一帧
    3. 节点只在时隙开始时发送帧
    4. 节点在时钟上是同步的二
    5. 如果两个或多个节点在一个时隙传输，所有站点都能检测到碰撞
2. 运行
    1. 当节点获取新的帧，在下一个时隙传输

    2. 传输时没有检测到碰撞，继续在下一时隙发送新帧
    3. 检测到碰撞：节点在每一个随后的时隙以概率p重传帧直到成功
3. 优点
    1. 节点可以以信道带宽全速连续运输
    2. 高度分布：仅需要节点之间在时隙上的同步
    3. 简单
4. 缺点
    1. 存在碰撞会浪费时隙
    2. 即使有帧要发送仍然可能存在空闲的时隙
    3. 节点检测碰撞的时间<帧传输的时间
    4. 需要时钟上同步
5. 效率:Np(1-p)^(N-1)，N无穷大极限为0.37

纯ALOHA(非时隙)

1. 简单、无需节点间在时间上同步
2. 当有帧需要传输：马上传输
3. 冲突的概率增加
4. 效率：p(1-p)^(2(N-1))，极限17.5%，效率更差

CSMA(载波侦听多路访问)

1. 在传输前先侦听信道
    1. 如果空闲，传输整个帧
    2. 忙，推迟
2. 碰撞检测(collision detection)：当一个传输节点在传输时一直在侦听此信道，检测到干扰则停止传输，等待一段随机时间
3. 碰撞仍可能发生
    1. 由信道传播延迟(channel propagation delay)造成:两个节点可能听不到正在进行的传输
4. 具有碰撞的CSMA(CSMA with Collision Detection,CSMA/CD)
    1. 适配器获取数据报，创建帧
    2. 发送前侦听信道CS：闲/忙
    3. 发送过程碰撞检测CD
    4. 发送方适配器检测到冲突，厨房其外还发送一个Jam信号，强化冲突：让所有站点都知道碰撞
    5. 如果放弃，适配器进入指数退避状态(二进制指数规避算法，exponential backoff)
5. 指数退避：适配器试图适应当前负载，在一个变化的碰撞窗口中随机选择时间点尝试重发
6. 效率:CSMA/CD效率定义：当有大量的活跃节点，且每个节点都有大量的帧要发送时，帧在信道中无碰撞地传输的爱不分时间在长期运行时间中所占的份额，efficiency=1/(1+5*dprop/dtrans),dprop表示信号能量在任意适配器之间传播所需的最大时间，dtrans表示传输一个最大长度的以太网帧的时间。比ALOHA更好

线缆接入网络：DOCSIS：TDM上行信道

1. 采用TDM的方式将上行信道分成若干微时隙：MAP指定
2. 站点采用分配给它的微时隙上行数据传输：分配
3. 在特殊的上行微时隙中，各站点请求上行微时隙：竞争

轮流 MAC协议(taking-turns protocol)

1. 信道划分MAC协议：
    1. 共享信道在高负载时是有效和公平的
    2. 在低负载时效率低下
2. 随机访问MAC协议
    1. 在低负载时效率高：单个节点可以完全利用信道全部带宽
    2. 高负载时：碰撞开销较大，效率极低，时间很多浪费在碰撞中

轮询：

1. 主节点邀请从节点依次传送
2. 主节点以循环的方式轮询(poll)每个节点
3. 优点
    1. 消除了困扰随机接入协议的碰撞和空时隙
4. 缺点
    1. 轮询开销：轮询本身消耗信道带宽
    2. 等待时间：每个节点需要等到主节点轮询后开始传输，即使只有一个节点，也许要等到轮询一周后才能够发送
    3. 单点故障：主节点失效时造成整个系统无法工作

令牌传递协议(token-passing)

1. 控制令牌(token)循环从一个节点到下一个节点传递
2. 令牌报文：特殊的一小段帧，在节点之间以某种固定的次序进行交换
3. 缺点：
    1. 令牌开销：本身消耗带宽
    2. 延迟：只有等到抓住令牌，才可传输
    3. 单点故障(token)：
        1. 令牌丢失系统级故障，整个系统无法传输
        2. 复杂机制重新生成令牌

交换局域网

1. 32bitIP地址：
    1. 网络层地址
    2. 前n-1跳：用于使数据报到达目的IP子网
    3. 最后一跳：到达子网中的目标节点
2. LAN(MAC/物理/以太网)地址：
    1. 用于使帧从一个网卡传递到与其物理连接的另一个网卡
    2. 48bit MAC地址固化在适配器的ROM，有时也可以通过软件设定
    3. 理论上全球任何2个网卡的MAC地址都不相同

网络地址和MAC地址分离

1. IP地址和MAC地址的作用不同
    1. IP地址是分层的
        1. 一个子网所有站点网络号一致，路由聚焦，减少路由表
            1. 需要一个网络中的站点地址网络号一致，如果捆绑需要定制网卡非常麻烦
        2. 希望网络层地址是配置的
    2. MAC地址是一个平面的

        1. 网卡在生产时不知道被用于哪个网络，因此给网卡一个唯一的标示，用于区分一个网络内部不同的网卡即可
        2. 可以往成一个物理网络内部的节点到节点的数据交付
    2. 分离好处
        1. 网卡坏了，ip不变，可以捆绑到另一个网卡的MAC上
        2. 物理网络还可以除IP之外支持其他网络层协议，链路协议为任意上层网络协议
    3. 捆绑问题
        1. 仅使用IP地址，那么仅支持IP协议
        2. 每一次接电都要重新写入网卡IP地址

LAN地址和ARP

1. 局域网上每个适配器都有一个唯一的LAN地址
2. MAC地址由IEEE管理和分配
3. 地址解析协议(Address Resolution Protocol,ARP):
    1. 在LAN上的每个IP节点都有一个ARP表
    2. ARP表：包括一些LAN节点IP/MAC地址的映射
    3. 寿命(TTL):地址映射失效的时间
    4. 路由在同一个LAN：路由器在自己的ARP表中，缓存IP-to-MAC地址映射关系，直到信息超时；ARP即插即用，无需网络管理员干预，节点自己创建ARP的表项
    5. 路由到其他LAN：先发送到网关路由器，路由器提取IP分组，交给上层IP协议，R创建帧，目标MAC地址设定

以太网(Ethernet)

1. 以太网帧结构
    1. 前同步码(preamble)：用来同步接收方和发送方的时钟速率
        1. 使得接收方将自己的时钟调到发送端的时钟
        2. 从而可以按照发送端的时钟来接受所发送的帧
    2. 目的地址(dest. address):包含目的适配器的MAC地址
    3. 源地址(source address):包含了传输该帧到局域网上的适配器的MAC地址
    4. 类型字段(type):允许以太网复用多种网络层协议
    5. 数据字段(data):承载了IP数据报
    6. CRC(循环冗余检测):使得接收适配器检测帧中是否引入了差错
2. 无连接、不可靠的服务
    1. 无连接：帧传输前，发送方和接收方间没有握手
    2. 不可靠：接收方适配器不发送ACK或NAK给发送方
    3. 以太网的MAC协议：采用二进制退避的CSMA/CD介质访问控制形式
3. 802.3以太网标准：链路和物理层
    1. 相同的MAC协议和帧结构
    2. 不同速率
    3. 不同物理层标准
    4. 不同物理层媒介
4. IEEE802.3z
    1. 使用标准以太网帧格式
    2. 允许点对点链路以及共享的广播信道
    3. 使用CSMA/CD共享广播信道
    4. 对于点对点信道高带宽全双工操作

5. 以太网使用CSMA/CD
    1. 没有时隙
    2. NIC如果侦听到其它NIC在发送就不发送：载波侦听
    3. 发送时，适配器当侦听到其它适配器在发送就放弃对当前帧的发送，冲突检测
    4. 冲突后尝试重传，重传前适配器等待一个随机时间
6. 以太网技术
    1. 转发器(repeater)：能够得到更长的运行距离，能再输入端接收信号并在输出端再生该信号

交换机

1. 过滤(flitering):决定一个帧应该转发到某个接口还是应当丢弃的交换机功能
2. 转发(forwarding):决定一个帧应该被导向哪个接口，并把该帧移动到那些接口的交换机功能
3. 交换机表中的一个表项包含
    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
4. 自学习(self-learning)
    1. 初始表为空
    2. 记录了发送节点所在的局域网网段
    3. 老化期(aging time)：经过老化期后，交换机接收到以该地址作为源地址的帧，就在表中删除这个地址。
5. 即插即用(plug-and-play device)
6. 交换机的性质
    1. 消除碰撞
    2. 异质的链路：链路彼此隔离，局域网的不同链路能够以不同的速率运行并且能够在不同的媒体上运行
    3. 管理：易于进行网络管理
7. 与路由器的比较
    1. 交换机
        1. 优点：即插即用，具有相对高的分组过滤和转发速率
        2. 缺点：交换网路的活跃拓扑限制为一棵生成树，ARP流量和处理量庞大，对于广播风暴并不提供任何保护措施
    2. 路由器
        1. 优点：分层次地网络寻址，分组不会通过路由器循环；分组不会被限制在一棵生成树上；并可以使用源和目的地之间的最佳路径；允许以丰富地拓扑结构构建因特网；对广播风暴提供了防火墙保护
        2. 缺点：不是即插即用的,需要人为地配置IP地址；路由器对每个分组的处理时间通常比交换机长

虚拟局域网(Virtual Local Network,VLAN)

1. 支持VLAN的交换机允许经一个单一的物理局域网基础设施定义多个虚拟局域网，在一个VLAN内的主机彼此通信。
2. 每个组构成一个VLAN，在每个VLAN中的端口形成一个广播域
3. VLAN干线连接(VLAN trunking):每台交换机上的一个特殊端口被配置为干线端口，以互联这两台VLAN交换机，干线端口属于所有VLAN
4. 802.1Q帧

1. 由标准以太网帧与加进首部的4字节VLAN标签(VLAN tag)组成，VLAN标签承载着该帧所属的VLAN标识符
2. VLAN 标签自身由一个2字节的标签协议标识符(Tag Protocol Identifier,TPID)字段、一个2字节的标签控制信息字段、和一个3bit优先权字段组成

链路虚拟化：网络作为链路层

多协议标签交换(Multiprotocol Label Switching,MPLS)

1. 目标：基于固定长度标签和虚电路的技术，在不放弃基于目的地IP数据报转发的基础设施的前提下，当可能时通过选择性地标识数据报并允许路由器基于固定长度的标签转发数据报来增强其功能。
2. 标签交换路由器(label-switched router):一个MPLS使能的路由器，一个MPLS加强的帧仅能在两个均为MPLS使能的路由器之间发送
    1. 基于标签的值进行分组的转发而并非检查IP地址
        1. MPLS转发表和IP转发表相互独立
    2. 弹性：MPLS转发决策可以和IP不同
        1. 采用源地址和目标地址来路由到达同一个目标的流，不同路径
        2. 如果链路失效，能够快速重新路由：预先计算好的备份链路(对于VoIP有效)
    3. 与IP路由相比：IP路由到达目标的路径仅仅取决于目标地址；MPLS路由到达目标的路由可以基于源和目标地址：快速重新路由：在链路失效时，采用预先计算好的路径
3. 概述：
    1. 建立基于标签的转发表-信令协议：支持逐跳和显式路由：路由信息传播，路由计算(基于Qos，基于策略的)，标签分发
        1. RSVP-TE：关注对MPLS信令所做的工作，采用该信令协议在下游路由器上来建立MPLS转发表
4. MPLS优点：
    1. 路由弹性：基于Qos，基于策略的
    2. 充分利用已有的硬件ATM，快速转发
    3. 支持流连工程，VPN
    4. 支持带宽等资源的分配

数据中心网络

负载均衡器：应用层路由

1. 接受外部的客户端请求
2. 将请求导入到数据中心内部
3. 返回结果给外部客户端

在交换机之间，机器阵列之间有丰富的互联措施

1. 在阵列之间增加吞吐
2. 通过冗余度增加可靠性

笔记本连接到互联网

1. 笔记本请求一个IP地址，第一跳路由器的IP地址，DNS的地址：采用DHCP
2. DHCP请求被封装在UDP中，封装在IP中，封装在802.3以太网帧中
3. 以太网的帧在LAN上广播，被正在运行的DHCP服务器接收到
4. DHCP服务器解封装IP分组，解封装UDP，解封装DHCP

5. DHCP服务器生成DHCP ACK包括客户端IP地址，第一条路由器的IP地址，DNS名字服务器地址
6. DHCP服务器封装，帧通过LAN转发，在客户端解封装
7. 客户端接收DHCP ACK应答

请求网站

1. DNS来查询网站IP地址
2. DNS查询被创建，封装在UDP段，封装在IP数据报中，封装在以太网帧中，帧传递给路由器，接口通过MAC地址，MAC地址通过ARP获取
3. ARP查询广播，被路由器接收，路由器用ARP应答，给出其IP地址某个端口的MAC地址
4. 客户直到第一条路由器MAC地址，可以发送DNS查询帧
5. 通过LAN交换机转发，包含DNS查询的IP数据报从客户端到第一条路由器
6. 通过一堆网络路由到DNS服务器
7. 被DNS服务器解封装
8. DNS服务器恢复给客户端网站的IP地址

HTTP请求

1. 客户打开到达Web服务器的TCP socket
2. TCP SYN段(第一次握手)域间路由到Web服务器
3. Web服务器用TCP SYNACK应答(第二次握手)
4. TCP连接建立
5. HTTP请求发送到TCPsocket中，IP数据报包含HTTP请求(第三次握手)最终路由到Web服务器
6. Web服务器用HTTP应答回应
7. IP数据报包含应答，最后被路由会客户端