**1.1 一些基本概念：**

路由器(server)、链路层交换机(link-layer switch)

分组经历的通信链路和分组交换机称为**路径**(route/path) 吞吐量（Throughput）是指在单位时间内成功传输的数据量

**网络协议**：因特网中信息的接受和发送，互联网主要协议为**TCP**/**IP**与因特网相连的接口为**套接字接口**(socket interface)

**协议**定义了两个或多个通信实体之间交换的报文格式和顺序，以及报文的发送/接受或其他时间产生的事件方面所采取的动作

原语：上层使用下层服务的方式

接入网（DSL）通过线缆接入，ICP将数据中心接入ISP，服务访问点：SAP传输层的服务访问点位port，上层的数据是sdu，加上本层的信息，变为pdu

**1.2 网络边缘(物理层)主机和服务器**

**主机**：**端系统**又称主机，用于容纳(运行)运用程序。主机又可以划分为**客户**和**服务器**。

很大的服务器还有**数据中心**(data center)。

**1.3 网络核心(网络层、链路层)**

端系统彼此交换**报文**.报文会被分为小的数据块,称为分组.源和目的地之间,每个分组都通过通信链路和**分组交换机**.交换机包括路由器和链路交换机。

共分为广域网(WAN)、城域网(MAN)、局域网(LAN)、个人区域网(PAN)

**1.3.1 存储转发传输与时延计算**

**存储转发传输**是指在交换机在开始向输出链路传输该分组的第一个比特之前，把必须接收到整个分组。分组交换包括虚电报和数据包。

通过条速率为(从队列中推出比特的速率)的链路组成的路径(**中间有个路由器**)，报文长为的时延为(传输时延)

更一般的，如果有组报文，那么传输时延会是

对于每条相连的线路，分组交换存在一个**输出缓存**。如果到达的分组需要传输到某条链路，但是该链路正忙于传输其他分组，该到达分组就必须在输出缓存中等待。这种时延被称为**排队时延**。由于缓存空间的大小是有限的，一些到达的分组或已经排队的分组会被丢弃，这被称为**分组丢失(包)**.

对于**电路交换**，每台主机和一个交换机直接相连。两台主机通信时，网络在两台主机之间建立一条专用**端到端连接**，**线路是预先预留好的。**

链路中电路通过频分复用(FDM)和时分复用(TDM)实现。

FDM线路是每条线路都连续得到部分带宽，频域上调整后耦合到一起传输。而TDM线路则是每条线路会在周期内一次性得到所有带宽。

电路交换时延：连接时延+发送时延+传播时延

发送时延=数据块长度/信道带宽

**1.4 分组交换网中的时延、丢包和吞吐量**

分组交换网会出现的时延：**节点处理时延**、**排队时延**、**传输时延**、**传播时延**。

**排队时延：**假设是分组到达的平均速率，传输速率为，分组有个比特，于是有流量强度。流量强度必须严格小于1否则排队延时极高.传输的第个分组的时延为

**传输时延:**见1.3.1

**传播时延:**由物理媒介决定。传播速度定义为，距离为，则传播时延为

节点的总时延为

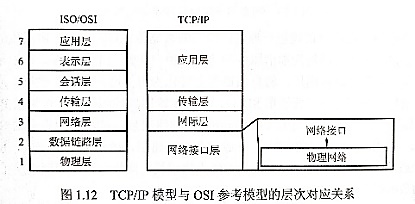
**端到端时延:**不考虑排队.这里*d*trans=*L*/*R*,于是

**1.5 协议层次与服务模型**

主要分为两种TCP/IP和OSI两种模型。

网络设计者以**分层**的方式来组织协议并实现协议的软件和硬件。某层向它的上一层提供**服务**.

TCP/IP与OSI:



应用层：交换信息，被称为报文(message)

运输层：报文段(segment)：TCP段，UDP数据报

网络层：分组packet（如果无连接方式：数据报datagram）

链路层:网络层的依靠。链路层的分组被称为帧

一层层包被称为封装

物理层：位(bit)

2. 应用层

**2.1 C/S模式**

**客户与服务器：**客户发起对话，服务器等待进程。主要特点：客户与服务器的地位不平等、客户机之间相互不通信、可扩展性不佳

**2.2 Web与HTTP**

**Web页面**是由对象组成的,一个对象**只是一个文件.**如果一个Web有一个HTML文本和五个JPEG图形，那么有6个对象。持久~可以发送多个对象，其中非流水需要等待相应再发送

GET获取数据POST上传数据PUT上传数据替换资源DELELTE删除资源HEAD：获取资源元数据，但是不返回实体内容OPTIONS：获取通信选项PATCH：部分更新指定资源

200 OK 301 Moved Permanently请求的对象已经被永久转移了；新的URL在响应报文的Location:首部行中指定客户端软件自动用新的URL去获取对象400 Bad Request一个通用的差错代码，表示该请求不能被服务器解读404 Not Found请求的文档在该服务上没有找到505 HTTP Version Not Supported

一个HTTP报文

Get /dir/p.html Host: [www.sjtu.edu.cn](http://www.sjtu.edu.cn)

Connection: User-agent:

Accept Language:

**Cookie**:一种站点对用户进行跟踪的方式。对用户标记后在用户端和服务器端各设置一个标号。

Web缓存：浏览器将所有http给缓存减少流量，提速ftp将控制连接21和数据连接20分开

**2.3 电子邮件**

一个电子邮件有3个主要的组成部分，**用户代理**、**邮件服务器**、**简单邮件传输协议**(STMP)

用户代理是用户与电子邮件系统的接口，用于阅读转发回复保存撰写邮件。邮件服务器是电子邮件的核心，用来发送和接受邮件。每个接收方在邮件服务器上有一个邮箱。邮件服务器必须能同时充当客户和服务器。端口25报文为7位ascii编码。邮件传输协议包括SMTP和POP3。其中SMTP是用户代理向邮件服务器发送邮件和邮件服务器之间发送邮件，而POP3（IMAP，HTTP）则是用户从服务器中获取邮箱中的邮件.**均基于TCP协议。**

邮件的报文格式如下：(区分SMTP命令)

From:a@163.com To:a@hotmail.com Subject: The Computer Network空行主体

**2.4 DNS：因特网的目录服务**

**主机名：**[www.sjtu.edu.cn](http://www.sjtu.edu.cn)

**IP地址：**114.145.191.180 **两者等价**

**域名系统**(DNS)是一种能进行主机名到IP地址转换的目录服务。其协议运行在UDP上，使用53号接口。

DNS的一些额外功能：主机别名、邮件服务器别名、负载分配。

**域名解析过程**

全世界拥有很多DNS服务器，包括根服务器、顶级域服务器、权威DNS服务器。

**根服务器：**负责提供顶级域服务器的IP地址。任何一个本地域名服务器无法解析时都要求助根服务器。

**顶级域服务器（tld）**：对于每个顶级域(.com,.edu)和国家顶级域(.cn,.uk,.jp)都有一个DNS服务器。

**权威DNS服务器**：具有公共可访问主机的每个组织机构都必须提供可被访问的DNS记录。

**本地DNS服务器**：每个因特网服务提供者(ISP)都有一个本地域名服务器。当主机与某个ISP连接时，ISP就会提供一个主机的IP地址。

**递归查询和迭代查询**

递归查询：本地主机所询问的本地域名服务器不知道被查询域名的IP地址，那本地域名服务器就以DNS服务器客户的身份，向根域名服务器继续发出报文请求IP地址。全是跟服务器做

迭代查询：当根服务器收到本地域名服务器发出的请求报文时，要给出查询的IP地址或告知本地域名服务器下一步应当访问哪个服务器，让该主机直接访问下一个IP。

**DNS缓存:**当某个DNS服务器收到一个DNS回答时会将回答映射到本地存储中，以便下次访问时直接使用。

一个存储在DNS服务器的报文(A:host,NS:域名,MX:mail,CNAME:别名)

(Name, Value, Type, TLL（时效）)

**2.5 P2P模式**

P2P模式中,主机之间义务对等.P2P模式是弱服务器模式,减轻了服务器的计算压力,客户机之间可以直接共享文档,可扩展性好,鲁棒性强.

**2.5.1 P2P结构的可扩展性**

定义作为服务器的上载速率，作为第对等方的上载速率，作为第对等方的下载速率。用表示分发的文件长度(以比特记)，表示要获得该文件副本的对等方数量。**分发时间**是所有个对等方获得该文件副本所需要的时间。

令表示具有最小下载速率的对等方的下载速率，也即

我们得到C/S模式下的最小分发时间

以及P2P模式的最小分发时间

**2.5.2 BitTorrent**

**最稀缺优先技术：**请求希望的和最稀缺的块。KaZaA有一个组长，Gnutella则是泛洪

DASH（多媒体流化服务，告示文件提供不同的块的url）CDN内容分发

***3. 运输层 进程间的逻辑通信***

应用需要传输层的数据丢失率/延迟/吞吐/安全

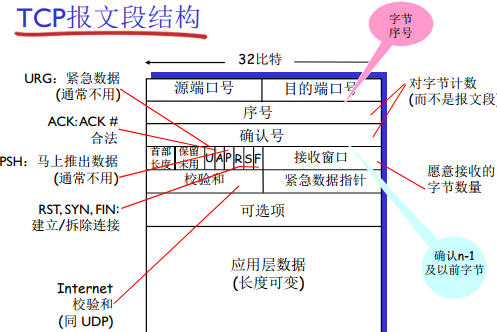
运输层是对不同主机上的应用之间进行的逻辑通信。运输层有两种协议：不可靠无连接的UDP、和可靠面向连接的TCP。**在生成socket时必须指明选择TCP还是UDP。**

与之相比，IP时不可靠服务。

**3.2 多路复用与多路分解**

**端口**(port)是能够让应用层的各种应用进程将其数据向下交付给应用层的过程。每个报文段都有特殊字段来指示该报文段要交付到的socket。端口号是一个16位的数，从0-65536，其中0-1023位是周知端口。**端口号区分进程**

|  |  |
| --- | --- |
| 32比特 | |
| 源端口号(source) | 目的端口号(dest) |
| 其他首部字段 | |
| 应用数据(报文) | |

****

**套接字**(socket)：(IP+端口)(TCP:4，UDP只有源）

**多路分解**(demultiplexing)：根据主机和端口将运输层报文段中的数据交付到正确的套接字。

**多路复用**(multiplexing)：在源主机从不同套接字中收集数据块，并位每个数据块封装上首部信息从而生成报文段。

**3.3 无连接运输：UDP**

UDP的优势：发送什么数据以及何时发送的应用层控制更为精细、无须连接建立、无连接状态、分组首部开销小。如果两个不同源IP地址/源端口号的数据报，但是有相同的目标IP地址和端口号，则被定位到相同的套接字

**3.3.1 UDP报文段结构**

|  |  |
| --- | --- |
| 32比特 | |
| 源端口号(source) | 目的端口号(dest) |
| 长度 | 校验和 |
| 应用数据(报文) | |

**3.3.2 UDP校验和**

校验和：16位报文加和，**如果溢出就加到后面，取反码得到校验和，检验时全1则对。**

UDP在**端到端基础上**在运输层提供差错检测

**3.4 可靠数据传输协议rdt**

**可靠数据传输协议**信道中传输比特不会损坏或丢失，而且所有数据都是按照发送顺序交付。

**3.4.1 一些可靠传输数据协议**

Rdt1.0:最可靠的通道。有传输端与接收端两个部分，传输端等待上层传资料进来，收到上面的资料后装成封包送出去。

Rdt2.0: 2.0考虑到了资料错误的情形，当接收端收到资料，会有ACK，NAK两种讯息，出现错误会传回NAK通知来源端再送一次(自动重传请求ARQ).以及停等(stop-and-wait)

Rdt2.1新增了sequence number，使用ACK与NAK来确认，封包的01号码可以用来确认是否重新传输封包停等协议：发送后等待应答

Rdt2.2移除NAK的讯息，在ACK中加入编号 就可以达到确认与否认的效果

rdt3.0 在传送端多了倒数计时器，封包如果超时仍未收到ACK或是收到不正确编号的ACK，则再送出封包一次。

**3.4.2 其他协议**

**停等协议和选择重传**

发送方利用率

为了避免利用率过低, 流水线：允许发送方在未得到对方确认的情况下一次发送多个分组

常见的流水线:回退N步GBN（接收窗口大小1，累计确认，只维护最老未确认分组的定时，最大尺寸2^n-1）,选择重传SR（选择性重传|单独确认，最大尺寸2^(n-1)）。

回退N步利用了滑动窗口(见TCP)

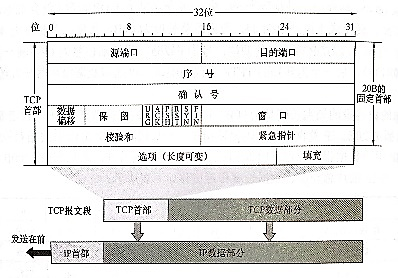
选择重传：重传可能出错的。

**3.5 面向连接的运输：TCP**

TCP提供的是全双工服务,收发两端都有缓存

Web服务器对每个连接客户端有不同的套接字，非持久链接对每个请求有不同的套接字

**3.5.2 报文**

****

**序号**：将字节流按顺序编号，序号指发送数据的第一个字节的编号

**确认号**:对方期望收到**下一报文段**第一个序号

**只确认流中第一个丢失为止的字节,**这是**累积确认 MSS：最大报文段大小**

**数据偏移**(首部长度):表明TCP首部长度，TCP首部长度范围为20-60字节

ACK确认 PSH推送(意味着缓存将满) RST复位(重传)FIN终止 URG紧急(使用紧急指针)

校验和:对源,目的,(保留字节(置0)、传输层协议号(TCP是6),TCP报文长(报头+数据))求和

**3.5.3 往返时间估计与超时**

估计往返时间:定义SampleRTT为某个时段测量的RTT, EstimatedRTT为估计的RTT值. EstimatedRTT的更新方法为(RTT忽略)

一般取0.125

DevRTT安全边界时间估算RTT偏差

一般取0.25

超时重传间隔TimeoutInterval定义为

超时间隔加倍

**超时重传最老的后沿报文段**

**冗余ACK**(duplicate ACK)就是再次确认某个报文段的ACK. 用于控制网络的丢包. 由于不能直接否认, TCP会对最后一个按序字节数据进行重复确认. 一旦收到3个就重传

**3.5.6 TCP连接的建立与释放**

**建立连接的三次握手解决半连接和接受老数据**

服务器进程出于LISTEN状态,等待连接请求

1. 客户的TCP向服务器TCP发送连接请求报文段,初始序号为*x*,SYN置1

2. 服务器TCP同意建立连接,为该TCP连接建立缓存和变量,SYN和ACK置于1,初始序号为*y，acknum=x+1*

3. 客户向服务器发送确认信号,ACK置1,

**连接释放的四次握手**

1. 客户的TCP向服务器TCP发送连接请求报文段, 初始序号为*u*, FIN置1

2. 服务器收到后即发送确认,

3. 服务器已无要发送的数据, 通知客户释放连接,FIN=1,ACK=1,

4. 客户机确认报文,随后确认释放连接,ACK=1,

**3.7 TCP拥塞控制**

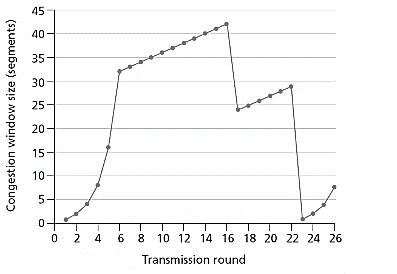
**输出端口N个流输出缓存：（RTT·C）/sqrt(N)**

**流量控制：应用接受的比tcp接受的慢，告知发送方，发送方限制未确认<=空闲bufffer**

**接受窗口**(recive window, rwnd)用于表示剩余的缓存数量,其计算方式为

**拥塞窗口**(congestion window, cwnd)用于向网络中发送流量的速率进行限制.我们有

rate 约等于congwin/RTT

TCP拥塞控制算法如下，超时大概率拥塞

3个重复ACK,快速重传

拥塞避免

超时重传

慢启动

Ssh=32 ssh=cwnd=24

慢启动Ssh=12

**三次重复ACK轻微拥塞，比超时好**

**1.慢启动与拥塞避免**

先发送一个MSS,首次确认后增大**一倍**cwnd（每个RTT）.同是指一个ssthresh(慢启动阈值),当cwnd> ssthresh时,停止使用慢开始而使用拥塞避免(每cwnd加1线性增长)

**2. 网络拥塞（TCP Tahoe）乘性减加性增**

当检测到拥塞时(未及时收到确认),就将ssthresh设置此时cwnd的一半,同时将cwnd置1,重新慢开始，阈值设为上次的一半。

**3. 快重传与快恢复（TCP Reno）**

快重传见3.5.3

快恢复是当收到三个冗余ACK时,发送方不会认为现在发生了严重的网络拥塞,ssthresh=cwnd/2,cwnd=ssthresh+3.**出现超时时仍同2处理**

一条连接的平均吞吐量定义为

其中指拥塞窗口大小.

4. 网络层:数据平面 转发和路由

网络层分为数据平面和控制平面

网络层功能：转发：将分组从路由器的输入接口转发到合适的输出接口口；路由：使用路由算法来决定分组从发送主机到目标接收主机的路径o路由选择算法o路由选择协议

**数据平面**：触及系统中的每个数据包/请求。负责服务发现、健康检查、路由、负载平衡、身份验证/授权和可观察性。传统转发基于目标地址+转发表，SDN（远程控制器与本地控制代理交互）基于字段+流表

**控制平面**:为网格所有正在运行的数据平面提供策略和配置.不触及系统的任何数据包/请求.控制平面将所有数据平面变成一个分布式系统，决定端到端路径

**网络服务模型:**确保交付、分组有序交付、确保最小带宽、安全性 **MTU最大传输单元**

**4.2 路由器工作原理**

路由器由输入端口、交换结构、输出端口、路由选择处理器组成。

路由的转发分为基于目的地转发和泛化转发

数据的交换分为经内存交换、经总线交换、将互连网络交换。

路由：运行路由选择算法／协议 (RIP, OSPF, BGP)-生成路由表

转发：从输入到输出链路交换数据报-根据路由表进行（最长前缀匹配）分组的转发

**输入排队：**如果两个输入队列的前端时发往一个输入队列的，那么其中一个必须等待

**输出排队：**如果缓存临近填满，那么需要丢弃一些分组。可能根据优先级或随机丢

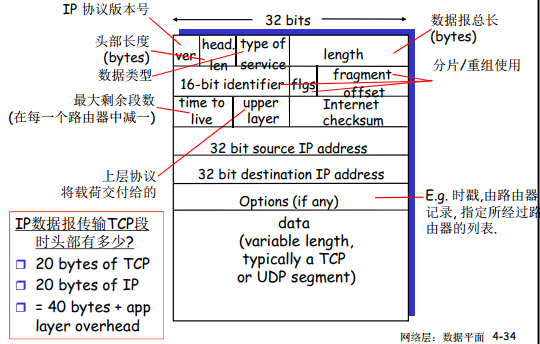
N个流缓存(C链路容量）：(RTT·C) / sqrt(N)

**一些分组调度：**先进先出（FIFO），优先权排队（非抢占式优先权排队）、轮转调度（一个类一起发出去）循环排队、加权公平排队（WFQ）每个类在同一时间间隔下可能获得不同的服务。每个类*i*会分配一个权*wi*,对于传输速率为*R*的链路，会获得的吞吐量。

**分片**：偏移量以8字节为单位，20字节头部，最后一个分片标记为0，前面MF=1，最后0

**4.3 IPv4/IPv6**

IPv4数据报，转发表是一个映射



**IPv4编址**

主机和物理链路的边界叫**接口**(interface), IP地址使用**点分十进制记法,32位，一个IP和一个接口关联**

为了产生隔离的网络，使用接口端隔离这些网络的端点，每个隔离的网络叫**子网**(subnet)

IPv4的子网划分采用**子网掩码**(network mask)的方式,区分子网利用了**最长前缀匹配.** **层次编址允许路由信息的有效广播**

A类地址:1.0.0.0-126.0.0.0 subnet:/8

B类地址:128.0.0.0-191.255.255.255 subnet:/16

C类地址:192.0.0.0-223.255.255.255 subnet:/24

D类地址:224.0.0.1-239.255.255.254 多点广播

E类地址:240.0.0.1-255.255.255.254 未启用

默认地址全0可以匹配所有，全1为广播地址，网络部分为0表示这个网络上的地址，127表示自环，网络+全1表示某网络广播

专用地址范围：Class A 10.0.0.0-10.255.255.255 MASK 255.0.0.0Class B 172.16.0.0-172.31.255.255 MASK 255.255.0.0 Class C 192.168.0.0-192.168.255.255 MASK255.255.255.0

网络前缀都相同的连续IP组成CIDR地址块

**动态主机配置协议**(DHCP).应用层，管理员能够配置DHCP,使得某个个主机每次与网络相连时得到一个相同的IP或被分配一个临时IP.又称作即插即用协议和零配置协议. DHCP 请求被封装在UDP段中,封装在IP数据报中，封装在以太网的帧中

分配IP的步骤为:

1. **DHCP服务器发现:** 客户发送DHCP发现报文,向UDP分组中端口67发送报文,传向广播地址255.255.255.255,自身源设置为0.0.0.0

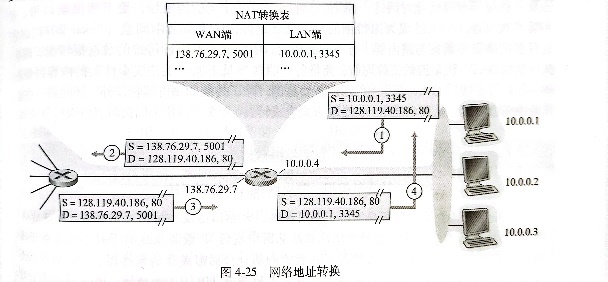
2. **DHCP服务器提供:** DHCP服务器发现DHCP报文时,会发送DHCP提供(offer)报文向客户做出相应.向所有节点广播,广播地址255. 255.255.255.报文包含事务ID,IP地址,网络掩码和地址租用期/

3. **DHCP请求** 客户从一个或多个服务器中选择一个,并向选中的服务器提供DHCP请求(request)报文相应

4. **DHCP ACK** 证实所使用的参数

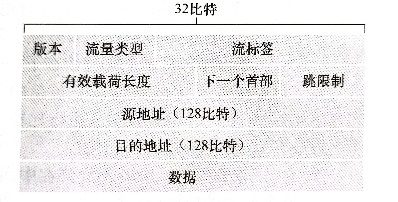
**网络地址转换(NAT) 安全**

有一些IP地址时被保留在**专用网络**(private network), 10.0.0.0/8, 172.16.0-31.0/16, 192.168. 0-255.0/24 转换过程如下图



利用NAT的IP,用不同端口区分LAN不同IP.

**IPv6数据报 报头60byte，地址128bit，每段16bit分为8端**

****

IPv4和IPv6利用隧道（用ipv4包起来）转换，头部 40字节，传输不分片，没校验和，::表示一系列连续的零

**4.4 泛化转发与SDN Software-Defined Networking**

软件定义网络(SDN)集中式的控制平面和分布式的数据平面.利用Openflow协议,将转发表下发给路由器.整个软件的功能只有收到分组、查找转发表、转发分组。控制平面-控制器+网络应用，计算和下发流表。

SDN的特性是基于流的转发、数据平面与控制平面分离、网络控制功能和可编程的网络。

管理容易，开放实现，可编程，水平集成，控制与数据平面分离。北向API与控制应用交互，南向API与交换机交互

1. 网络层:控制层面

确定数据报从源端到目的端如何选择路由

**在一个网络中路由器主机的通信链路层解决**

**路由选择应该正确简单公平健壮稳定最优**

**5.2 路由选择算法（RIP）（UDP协议）**

**链路状态路由算法(LS)收敛快，健壮:**1.发现相邻节点，获知对方地址2.测量到相邻的代价3.组装分组描述相邻节点的情况4.扩散分组（标记年龄，需要发送的和需要ack的）5.利用Dijkstra算法(D(v)=min{D(v),D(u)+c(u,v)},每次加最小nlogn

这种算法的特点在于路由器只知道周围的部分信息，只有当链路状态发生变化时，路由器才向所有路由器发布。**是全局算法，**在运行Dijkstra前需要获取全局信息。

**距离向量路由选择算法**

**当x从邻居收到DV时，自己运算，更新它自己的距离矢量，如果邻居变化则通告，复杂度低**

**DV算法是一种迭代的,异步的,分布式算法.**这种算法与Bellman Ford算法一致,转移方程为

**DV**算法可能出现的问题：好消息快，坏消息慢。链路开销改变和链路故障（选择黑洞，改变开销后反复横跳），无穷计数问题。

解决方案：**毒性逆转**：改变开销时所有涉及的路径时通报为无穷。例如，当的开销变为60时，选择路线，会被告知为无穷。**但这种解决方案不完备。**

**水平分裂：对下真距离，对上给无穷**

**两者比较：**DV算法报文复杂性低，收敛慢，由于是分布式算法一个错误会发展到整个网络，鲁棒性低。LS算法报文复杂性高，但是收敛快，路的计算彼此分离，鲁棒性高。

**内部网关协议：RIP，OSPF Open Shortest Path First**

**5.3 OSPF算法（IP协议）**

核心:路由器自治,本自治系统向所有路由器发送信息（洪泛法）;发送的信息说与本路由器相邻的所有路由器状态;只有当链路状态发生变化时,路由器采用洪泛法向所有路由器发送信息;OSPF时网络层协议,直接利用IP数据报.

**工作原理：已知全网拓扑图后，直接利用Dijkstra计算到各自网络的最优途径。**

OSPF的优点：安全、可使用多条路径、对单播和多播路由的支持。支持在单个自治系统，有多重代价矩阵，允许有多个代价相同的路径存在。

平面路由：规模巨大，不便于管理。

层次路由：网关路由器：AS边缘路由器，可以连接到其他AS

**5.4 ISP之间的路由选择：BGP（TCP协议）**

两个路由之间的选择协议称为边界网关协议（BGP）。其目的为从邻居自治系统（AS）（多个网络的集合）获得前缀的可达性信息以及前缀的最好路由。基于DV算法记录路径来避免环路。对于每个AS，每台路由器要么是一台网关路由器，要么是一台内部路由器。BGP发送的报文会同时告知发送的路径以及后缀。

每条直接连接以及所有通过该链接发送的. BGP报文被称为**BGP连接.**跨越两个AS的BGP连接被称为外部BGP(eBGP)连接,在相同AS中两个路由器连接被称为内部BGP(iBGP).

**如何确定最好的路由 Inter-AS更关注策略**

**热土豆路由选择**

原则：自私算法，减小在自己AS内的开销

步骤如下：

1. 从AS间协议学到经多个网关可达子网*x*

2. 使用来自AS内部协议的路由选择信息，以决定到达每个网关的最低开销路径的开销

3. **热土豆路由选择：**选择具有最小最低开销的网关

4. 从转发表确定通往最低网关的接口*I*.在转发表中加入表项(*x,I*)

**路由器选择算法：**

1.路由被指派一个本地偏好，选偏好最高的

2.从余下相同偏号路由中选择最短路径

3.再余下相同路由和路径选用热土豆路由选择

4.在剩下的的用BGP标识符

**路由器补充：**路由器的任务是连接不同网络并完成路由转发。路由表包含目的网络IP地址、子网掩码、下一跳IP地址，接口。

**ICMP:** 因特网报文控制协议,用于彼此沟通网络层的信息,包括确认链接是否不可达,TTL过期等。ICMP位于网络层但是通过ip传递

Traceroute 源主机发送逐步增加TTL的UDP段，接收中间路由器的ICMP报文（类型11，代码0）和目的地的ICMP端口不可达（类型3，代码3）来记录RTT并确定路径，一个ntr3次

6. 链路层与局域网

数据链路层通常分为两个子层：MAC（Media Access Control，介质访问控制）子层和LLC子层。LLC子层主要负责提供逻辑链路控制功能，包括服务访问点，流量控制，差错控制，帧的封装解封装；MAC子层：负责与物理介质的直接交互，包括介质访问控制、帧的传输和接收等。

在链路层中任何设备均称为**节点**,与通信信道被称为**链路**,传输的数据被封装到**链路层帧**中.半全工可以双向传输但一次只能一个。介质访问控制协议（MAC，Media Access Control）是数据链路层（OSI模型的第二层）的一部分，负责控制网络设备如何在共享的通信介质上进行数据传输，包括csma等。

**6.1 链路层提供的服务:**成帧,链路接入,可靠交付,差错检验和纠正

**6.2 差错检验和纠错技术**

即使检测差错,也有可能未检测出的比特差错

**6.2.1 奇偶校验**

假设 发送了*d*个比特,那么再加一位构成*d*+1个比特流.在**偶校验**中生成的比特流出现了奇数个**1**,或者**奇校验**中生成的比特流出现了偶数个**1**.这种未检测出差错的概率是50%.

使用**二维奇偶校验:**将*d*位比特流划分位*i*行*j*列,每行和每列增加一行和一列.然后每行和每列进行补位.

该做法出现单个比特的差错课以纠正.校验行和列也能被检测**但不能纠正.**

**6.2.2 因特网检验和**

对接受的数据(含校验位)求和并取反码,如果发现**不是全1就说明出现了差错**

**6.2.3 循环冗余检测**

现在计算机网络广泛使用的差错检测技术基于

考虑*d*比特的数据*D*,发送节点要和接收节点协商一个*r*+1位的比特模式,称之为*G*. **我们要求**在*D*后附加一个*r*位的CRC比特*R*,使得恰好能被*G*整除.

r为生成多项式最高次

*R*的计算方法为

也即*D*左移*r*位除以（异或）*G*取余(补到r位)

CRC能够1、双、长度等于或小于r位的错误，长度为r+1的检查不出概率1/(2^(r-1))，大于r+1的突发错误检查不出的概率为1/(2^r)。

**6.3 多路访问链路**

**信道划分/随机访问/依次轮流**

**点对点链路**是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.**广播链路**能够让发送和接受节点都连接到相同的、单一的、共享的广播信道上。如何协调多个发送和接收节点对一个共享广播信道的访问就是**多路访问问题**

规范在共享广播信道上的传输行为。

我们希望多路访问协议是简单的、去中心化的

**6.3.1 信道划分协议**

TDM将时间划为时间帧,并进一步划分每个时间帧为*N*个时隙。每个节点的专属传输速率为*R*/*N*.缺点在于一个节点只能受限于传输速率的限制，即使只有这一个节点要发送内容。

FDM是划分频域，与之类似。

还有一种是CDMA协议，见第七章

**6.3.2 随机接入协议**

核心：一个节点总是以全部速率发送，当由碰撞时，涉及碰撞每个节点反复重发它的帧，指导帧能够无碰撞的通过。**或者在重发该帧前等待一个随机时延。**

**1. ALOHA**

纯ALOHA协议，当发送数据时，不检测直接发送，如果发现出现碰撞，那么他就会在碰撞帧传输完后立即以概率*p*重传该帧,否则等待一帧保持空闲，下一帧继续以*p*等待。

**效率分析：**假设某个节点要成功传输，那么就要保证在这个间隔前面的一个间隔其他节点不传输，和自己所在的帧能够传输，其他节点不传输，于是传输成功的概率为,最好效率为1/(2e)

**2. 时隙ALOHA**

时隙ALOHA要做如下假设：

1. 所有帧都由*L*比特组成。

2. 时间被划分成*L*/*R*的时隙(传一帧的时间)

3. 节点只在传输起点传输帧

4. 节点是同步的

5. 如果出现碰撞,时隙结束所有节点都会发现

时隙ALOHA的协议为：

1. 当有一个新帧要传输时,到下一个时隙传输

2. 如果没有碰撞,成功传输

3. 如果有碰撞,时隙结束时检测到.接下来每个时隙都有*p*的概率重传该帧,直到成功

**效率分析:**一个节点发送成功的概率为,*N*个节点为,最好效率为1/e=37%,以及37%空闲,26%碰撞

**3. 载波侦听多路访问(CSMA)**

**原理:**

1. 传输时保证其他人不再传输(载波侦听)

2. 有传输干扰就停止传输(碰撞检测)

**问题:**由于时延,使得在侦听到空闲时可能已经有人在传播了,**信道传播时延**起到了决定作用

**4. 具有碰撞检测的载波侦听CSMA/CD)有线**

其工作流程可简单概括为:

1. 从网络层获取数据报

2. (先听后发)如果检测到空闲,传输

3. (边听边发)传输时监视其他节点流量

4. (冲突停发)如果无冲突就发送,有冲突中止.

5. (随机重发)随机等待时间,重复2

碰撞k次从{0~2^k-1}选一个r，经过r\*512位时后重传 争议期：信号在最远两个端点之间往返传输的时间

**效率:**定义信号在两个适配器之间传输的最大时间*d*prop,传输一个最大长度以太网帧的时间为*d*trans,有效率:1/()

**CSMA/CA**

发送方

1如果站点侦测到信道空闲持续DIFS长，则

传输整个帧 (no CD)2如果侦测到信道忙碌，那么 选择一个随机回退值，并在信道空闲时递减该值；如果信道忙碌，回退值不会变化到数到0时（只生在信道闲时）发送整个帧如果没有收到ACK,增加回退值，重复2

接收方如果帧正确，则在SIFS后发送ACK

冲突避免：发送方首先使用CSMA向BS发送一个小的RTS分组RTS可能会冲突（但是由于比较短，浪费信道较少）BS广播clear-to-send CTS，作为RTS的响应CTS能够被所有涉及到的节点听到发送方发送数据帧

其它节点抑制发送

**6.3.3 轮流协议**

我们希望当*M*个节点活跃时,每个节点的吞吐量接近*R*/*M*.

**1. 轮询协议:**某节点被指定为主节点,主节点以轮询的方式告诉每个节点能最多传输帧的数量. 缺点在于引入了轮询时延,鲁棒性差

**2.** 令牌(token)传递协议.没有主节点,一个令牌在节点之间交换.只有当有帧发送时才会持有令牌,否则立刻转交令牌.缺点:某个节点崩溃会使整个传输过程崩溃

MAC协议总结多点接入问题：对于一个共享型介质，各个节点如何协调对它的访问和使用?

信道划分：按时间、频率或者编码• TDMA、FDMA、CDMA随机访问 (动态)• ALOHA, S-ALOHA, CSMA, CSMA/CD•载波侦听:在有线介质上很容易,但在无线介质上比较困难• CSMA/CD：802.3 Ethernet网中使用• CSMA/CA：WLAN中使用依次轮流协议•集中：由一个中心节点轮询；分布：通过令牌控制•蓝牙、FDDI、令牌环

**6.4 交换局域网**

常见拓扑:星形,环形,总线形,星+总

**6.4.1 链路层寻址与ARP**

**MAC地址(或LAN地址,物理地址)**是链路层上的地址.一个MAC地址有6字节,有248种可能.由12个16进制数组成,如1A-2B-3C-4D-5E-6F.**每个硬件的MAC地址唯一**

MAC广播地址:FF-FF-FF-FF-FF-FF

**地址解析协议**(ARP: Address Resolution Protocol)是网络层地址和链路层地址之间的转换.每台主机或路由器在内存中有一个**ARP表**,包含IP地址与MAC地址的映射关系,以及一个寿命(TTL)值用于确认映射时长

当ARP表无目标主机表项时,它会构造一个包含发送和接受方IP和MAC地址的ARP表项,然后询问子网上所有路由器和主机找到地址.

**ARP是在网络层和链路层之间的协议**

**发送数据报到子网外:利用以太网ARP**

**6.4.2 以太网 无连接，不可靠**

20世纪90年代人们利用**集线器**转发数据,21世纪初改用**交换机**,避免了碰撞.

**以太网的帧结构为**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 前同步码 | 目的  地址 | 源地址 | 类型 | 数据 | … | CRC |

**数据字段(46-1500字节)**承载了IP数据报.若IP数据报不足46字节,需要填充

**目的地址(6字节)**目标MAC

**源地址(6字节)**源MAC

**类型字段(2字节)**使用协议编号

**CRC(4字节)循环冗余检测**

**前同步码(8字节)**前7个字节为10101010,最后一个为10101011,用于激活适配器

以太网向网络层提供无连接和不可靠服务

**6.4.3 链路层交换机**

交换机的特性在于对子网中的主机和路由器都是**透明的,**接口留有缓存

**1. 转发与过滤**

借助交换机表完成这两项功能.如果表中有该地址**直接转发,**如果没有,进行广播.

交换机表的地址也包括**存放在表中的时间**

**2. 自学习:**接收到入帧记录接口,一段时间(老化期)没接到帧就删掉这个地址

**3. 性质:**消除碰撞, 异质的链路, 管理

**6.4.4 VLAN**

交换局域网缺乏流量隔离,浪费交换机,很难管理用户.虚拟局域网解决了这一问题.

VLAN可用于分割动议子网,也可以配置VLAN接口来连接不同子网(方式为将两个交换机的接口与一台外部路由器相连).

更具拓展性的做法为**VLAN干线链接,**由IEEE802.1Q外加VLAN标签组成.

VLAN也可以基于网络层协议(IPv4,IPv6)和其他准则进行定义

流量隔离: 从/到1-8端口的流量只会涉及到1-8也可以基于MAC地址进行VLAN定义；动态成员: 成员可以在VLANs之间动态分配router；在VLANs间转发:通过路由器进行

转发 (就像他们通过各自的交换机相联一样)

MPLS网络按照标签label进行分组的转发，类似于VC；有基于标签的转发表；基于虚电路表，IP vs线路交换。标签交换的过程：入口路由器:LER对进入的分组按照EFC的定义打上标签；在MPLS网络中（虚拟成了链路）对分组按照标签进行交换；到了出口路由器，再将标签摘除。支持MPLS的路由器组构成的网络，从IP网络的角度来看虚拟成了链路 标签封装：一些列标准定义了在ATM,FR和以太网中如何封装，利用原有网络中的机制VCI，或者定义新的标签

**一个打开PC网页的过程(chap6 p31)**

PC首先使用DHCP获取IP地址.PC首先在DHCP服务器发现步骤中创建一个指定为255.255.255.255的特殊IP数据报,并将其放在以太网帧中并在以太网中广播,然后,按照DHCP协议中的步骤,PC能够获得具有给定租约时间的IP地址。

以太网上的DHCP服务器为PC提供了第一跳路由器的IP地址列表、PC所在子网的子网掩码以及本地DNS服务器地址(如果存在的话).

由于PC的ARP缓存最初是空的，PC将使用ARP协议来获取第一跳路由器和本地DNS服务器的MAC地址**。**

PC将首先获得想要下载的网页的IP地址.如果本地DNS服务器没有IP地址,则PC将使用DNS协议查找网页的IP地址。

一旦PC拥有了网页的IP地址,如果网页不驻留在本地Web服务器中,它将通过第一跳路由器发送HTTP请求.HTTP请求消息将被分割并封装到TCP数据包中,然后进一步封装到IP数据包中,最后封装到以太网帧中,PC将以太网帧发送到第一跳路由器.一旦路由器接收到这些帧,就会将它们传递到IP层,检查其路由表然后将数据包从所有接口发送到正确的接口.

然后,IP包将通过互联网路由,到达Web服务器

承载网页的服务器将通过HTTP响应消息将网页发送回PC.这些消息将被封装到TCP数据包中,然后进一步封装到IP数据包中.这些IP数据包遵循IP路由,最终到达第一跳路由器,然后路由器将这些IP数据包封装到以太网帧中,将它们转发到PC上.

7. 无线网络与移动网络

无线网络具备以下要素:无线主机,无线链路,基站,网络基础设施,单跳(有无基础设施),多跳(有无基础设施)

**7.2 无线链路和网络传播**

无线链路特征:递减的信号强度,来自其他源的干扰,多路径传播.

**信噪比:**所受到的信号与噪声强度的相对测量

*C=B*log2(1+*S*/*N*)

**比特差错率:**出现错误比特概率,信噪比负相关

可能出现的问题:隐藏终端问题,

CDMA(码分多址):对每个比特都要乘一个信号的比特进行编码,这个信号的变化速率(**码片速率**)比较高.CDMA编码将每个比特时隙细分为*M*个微时隙,编码由*M*个1或-1的比特流组成,

对于第*i*个数据比特*di*的第*m*个微时隙,CDMA编码器输出的*Zi,m*=*di·cm*.对于没有干扰发送方,恢复初始数据比特流

利用这种方式,即使收到了很多个不同的信号,也能通过乘*cm*快速复原

**7.3 WiFi: 802.11无线局域网**

WiFi的标准为IEEE802.11无线局域网

**体系结构:**基本模块:基本服务集(BSS),包括一个或多个无线站点以及作为**接入点(AP)**的中央基站.部署AP的无线局域网被称为**基础设施无线局域网.**

WiFi丛林是指在一个物理位置中,能同时收到很多人的AP时,你需要主动与 一个IP相连.

扫描和监听信标帧被称为被动扫描,无线主机也可以主动扫描

**802.11 MAC协议: 带碰撞避免的CSMA协议(CSMA/CA)**

首先进行链路层确认,先发一个校验帧等待其回复,然后等待一个短帧间隔.

1. 如果监听到空闲,会在短时间内发送该帧

2. 否则,随机选取一个回退值,如果空闲递减,如果忙不变

3. 计数值减到0, 发送整个数据帧

4. 如果收到确认,要发送下一帧的话执行第2步,未收到进入第2步重新回退

**高级特色:速率适应,功率管理**

**蓝牙:**工作距离短,功耗低成本低,被称为无线个人区域网络(WPAN)或者微微网

蓝牙有79个信道,在信道伪随机在一个时隙内发生改变

蓝牙必须是自组织的,存在邻居发现问题.

一旦发现潜在用户,就会执行蓝牙寻呼

**7.4 蜂窝网络 4G, 5G**

**核心设备:** 移动设备,基站,归属用户服务器,网络路由器,管理实体(对其执行身份验证,路径设置,小区位置跟踪)

LTE协议栈:链路层分为三个子层:**分组数据汇聚,无线链路控制,介质访问控制(MAC)**

**附加功能:** 网络连接(连接到基站,相互鉴别,移动设备到PDN的路径配置),睡眠模式

**5G**:分为450MHz-6GHz和24GHz-52GHz

5G增加的标准:eMBB(增强型移动带宽),URL LC(极可靠低时延通信),mMTC(大规模机器类通信). 增加的核心功能:用户平面功能.接入和移动性管理功能.会话管理功能

1. 计算机网络中的安全

安全性保护免受未经授权的访问、数据泄露和攻击。

安全通信具有以下性质:机密性,报文完整性,端点鉴别,运行安全性

入侵者可能会窃听或修改(插入/删除)报文内容

**8.2 密码学原理**

Alice将明文使用加密算法生成密文. 同时提供密钥, Bob利用解密算法还原密文. **对称密钥系统中**二者密钥相同且秘密,**公开密钥系统**使用一对密钥,一个两人知道,一个只有一人知道

**8.2.1 对称密钥**

**凯撒密码**: 将某个字母用字母表后*k*个字母替换. 改进方法为**单码代替密码**,也即用一个字典对字母进行替换

入侵者可能采用**唯密文攻击**(对密文强行解密),**已知明文攻击**(明文和密文都知道),**选择明文攻击**(只知道一部分密文对应的明文)

**多码代替密码:**有两个凯撒密码,使用两种不同的密码本依照特定的顺序交替生成

**块密码:**将每个报文进行分块,每个块利用给定的字母表一一映射

**密码块连接:**在第一个报文发送一个随机值,然后发送和接受使用编码块代替的随机数

**8.2.2 公开密钥加密**

发送者使用公开密钥进行加密,而接收者使用私钥和特定算法进行解密

**RSA:** 选择两个大素数*p*,*q*,计算*n=pq*,*z=*(*p*-1) (*q*-1).选择一个小于*n*的数*e*,与*z*互质,找到一个*d*使得*ed*-1可被*z*整除.公钥为(*n,e*),私钥为(*n,d*)

Alice的加密为*c=me*mod *n*

Bob的解密为*m=cd*mod *n*

已知*p,q*容易得到*d*,安全依赖于因数分解复杂

**8.3 报文完整性与数字签名**

**密码散列函数(Hash):***x*≠*y,H*(*x*)=*H*(*y*)找不到

**常见函数:**MD4 SHA1

**报文鉴别码:**生成报文*m*,与秘密*s*级联生成*m+s*

*H*(*m+s*)即为鉴别码附在*m*后.Bob知道*s*,可以算出*H*(*m+s*),Bob计算与Alice传的一致则正常

**数字签名:**Bob用私钥签署文档,Alice利用公钥计算出初始文档.**只有知道私钥才能签署文档**

**公钥认证:**将公钥和认证实体利用认证中心(CA)来完成

**8.4 端点鉴别:计算机网络向另一个实体证明其身份的过程**

**8.5安全电子邮件:机密性,发送方鉴别,完整性**

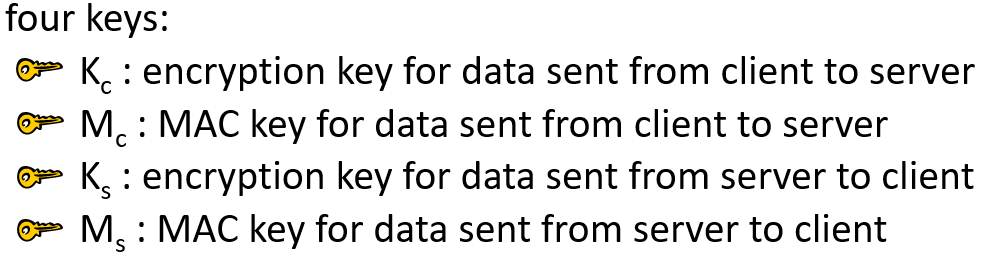
Alice哈希签名,级联后发送,Bob验证签名哈希

电子邮件加密范例:PGP

**8.6 使TCP更安全:TLS(运输层安全性)**

具体步骤:握手,密钥导出,数据传输,TLS记录

反截断攻击:加一个字段确认是否终止



**8.7 网络层安全性 IPsec与虚拟专用网(VPN)**

一些部门需要成立专用网络

协议族为:鉴别首部,封装安全性载荷

源和目的创建的网络连接称为**安全关联**

**IPSec有隧道模式和运输模式两种分组**