目录

[数据链路层 2](#_Toc8973904)

[1、使用点对点信道的数据链路层 2](#_Toc8973905)

[2、点对点协议PPP 3](#_Toc8973906)

[3、使用广播信道的数据链路层 4](#_Toc8973907)

[网络层 7](#_Toc8973908)

[1、网际协议IP 7](#_Toc8973909)

[2、划分子网和构成超网 8](#_Toc8973910)

[3、网际控制报文协议ICMP 9](#_Toc8973911)

[4、互联网路由选择协议 10](#_Toc8973912)

[5、IPv6 14](#_Toc8973913)

[6、IP多播 15](#_Toc8973914)

[运输层 18](#_Toc8973915)

[1、运输层协议概述 18](#_Toc8973916)

[2、可靠传输的原理 19](#_Toc8973917)

[3、TCP报文的格式 20](#_Toc8973918)

[4、TCP可靠传输的实现 20](#_Toc8973919)

[5、流量控制（拥塞控制） 20](#_Toc8973920)

[6、TCP连接管理 22](#_Toc8973921)

[应用层 24](#_Toc8973922)

[1、域名系统 24](#_Toc8973923)

[2、文本传输协议 25](#_Toc8973924)

[3、万维网WWW 27](#_Toc8973925)

# 数据链路层

主要内容：

a、数据链路层点对点信道和广播信道的特点，以及这两种信道所用的协议（PPP协议和CSMA/CD协议）的特点。

b、数据链路层的三个基本问题：封装成帧、透明传输、差错检测。

c、以太网MAC层的硬件地址。

d、适配器、转发器、集线器、网桥、以太网交换机的作用以及使用场合。

## 1、使用点对点信道的数据链路层

* 点对点信道的数据链路层在进行通信时的主要步骤（A对B通信）：

a、结点A的数据链路层将网络层交下来的IP数据报添加首部和尾部封装成帧

b、结点A将封装好的帧发送给结点B的数据链路层

c、若结点B的数据链路层收到的帧无差错，则从收到的帧中提取出IP数据报上交给上面的网络层；否则丢弃这个帧。

* 数据链路层的三个基本问题：封装成帧、透明传输、差错检测

1. 封装成帧

一帧包含帧首部、尾部以及数据部分。首部和尾部一个重要作用就是进行帧定界。此外，首部和尾部还包含很多控制信息。为了提高传说效率，按理来说，数据部分越大越好，可是，每一个链路层协议都规定了所能传送帧的数据部分长度上限（MTU）。

1. 透明传输

数据链路层的透明传输是指无论什么样的比特组合的数据，都能原封不动，毫无阻碍地通过数据链路层。因此，对于传送的数据来说，这些数据“看不见”有什么能妨碍数据传输的东西。

可是，在某些编码中，数据部分却可能出现帧定界符的编码，这样一来，就会误以为数据到此处就停止了，而把剩下的部分丢弃。为了解决透明传输的问题，可用字符填充的方法（也就是在数据部分在帧定界符前加个转义字符，若数据部分可能出现转义字符，则在该转义字符前加个转移字符）

1. 差错检测

通过循环冗余检验CRC，在每一帧的数据后面加循环冗余码。

## 2、点对点协议PPP

是一种简单的数据链路层可靠传输协议。

1. 主要特点

* **简单**接收方每收到一个帧，就进行CRC检验。如果CRC检验正确，就收下这帧，反之就丢弃，其他的什么也不做。
* **封装成帧**PPP协议必须规定特殊的字符作为帧定界符
* **多种网络层协议**PPP协议必须能够在同一条物理链路上同时支持多种网络层协议（如IP和IPX等）的运行。当点对点链路所连接的是局域网或路由器，PPP协议也必须支持链路所在的局域网和路由器上运行的协议。
* **多种类型链路**PPP必须支持在多种类型的链路上运行，如串行的和并行的，同步的和异步的，低速的和高速的。
* **差错检测**PPP协议要能检测有差错的帧并丢弃。
* **检测连接状态**必须具有一种机制能及时自动检测链路是否正常工作。
* **最大传送单元**PPP协议必须对每一种类型的点对点链路设置最大传送单元MTU的标准默认值。若高层协议发送的数据长度超过MTU的值，就丢弃这样的帧并报错。MTU是数据链路层可传送帧的数据部分的最大长度，而不是帧总长。
* **网络层地址协商**PPP协议必须提供一种机制使通信的两个网络层（如两个IP层）的实体能够通过协商知道或配置彼此的网络层地址。
* **数据压缩协商**必须提供一种方法来协商使用数据压缩算法。

1. 组成

PPP协议有三部分组成：

* 一种将IP数据报封装进串行链路的方法
* 一个用来建立、配置和测试数据连接的链路控制协议LCP
* 一套网络控制协议NCP

1. 帧格式

* 各字段意义

首部四个字段，尾部两个字段。

首部字段：F（1字节，标志字段，规定为0x7E），A（1字节，首部地址字段，至今未使用），C（1字节，控制字段，至今未使用），协议字段（2字节，若为0x0021，则信息部分是IP数据报；若为0xC021，则是链路控制协议LCP的数据；若为0x8021，则是网络层的控制数据。）

尾部字段：FCS（2字节，CRC检验序列），F（与首部同）

* 字节填充

若数据中出现了标志字符，就对其进行转义。

* 零比特填充

若数据部分出现6个连续的1（标志字段），就在第五个1后添加一个0，也是为了实现透明传输。

## 3、使用广播信道的数据链路层

1. 共享信道

共享信道在技术上有两种方法：

* **静态划分信道**如频分复用，时分复用、波分复用和码分复用。这种信道划分方法代价较高。
* **动态媒体接入控制**又称多点接入，其特点是信道并非在用户通信时固定分配给用户。这里分以下两类：

a、随机接入所有用户可随机地发送信息。但如果两个或多个用户同时发送信息，就会产生碰撞，因此，必须有解决碰撞的协议。

b、受控接入用户不能随机地发送信息，而是必须服从一定的控制。

**注：**受控接入一般用得少

1. 以太网的两个标准

局域网的数据链路层被拆成两个子层，即逻辑链路控制LLC和媒体接入控制MAC子层。

1. CSMA/CD协议

最早的以太网将许多计算机都连接到一根总线上。在每一台计算机发送信息时，要在发送帧首部写明目的计算机地址。所有计算机都可以收到该帧，可是只有硬件地址是目的地址的才会接收。为了通信方便，以太网采取以下两种措施：

a、采用无连接方式。对有差错的帧（会丢弃）是否重传由高层决定，若高层是TCP协议，就会要求重传。当有一台计算机发送信息时，其他计算机不能发信息，为了避免两台计算机同时通信，它使用的协议是CSMA/CD（载波监听多点接入/碰撞检测）

b、以太网发送的数据都用曼彻斯特编码。

* CSMA/CD协议的特点

**多点接入**说明这是总线型网络。

**载波监听**就是检测总线上有没有其他计算机在发送信息

**碰撞检测**就是边发送边监听，若检测到总线上发生碰撞，就立刻停止发送。

使用CSMA/CD协议时，不可能进行全双工通信，而只能半双工。每个站发送信息后，在经过一个端到端的往返时延的时间后，才能保证自己的通信没有发生碰撞（如果无碰撞的话）。若再次期间检测到碰撞，就要重传，一般采用截断二进制指数退避算法。

**最短帧**考虑这样一种情况，某个站发送很短的一帧，在发送完毕前并未检测到碰撞，但在传送时发生碰撞，这时发送方不知道。因此，规定一个最短帧64字节（512bit，即发送时间刚好是一个往返延迟）。因此，收到的小于64字节的都是因发送冲突而无效的帧。

**碰撞强化**即检测到碰撞时，再继续发送一个32或64比特的人为干扰信号。

**帧间最小间隔**为了使刚刚收到的数据帧来得及清理，做好接收下一帧的准备而设置的帧长

1. 以太网的MAC层

**MAC地址**局域网上的每台计算机固化在适配器ROM上的地址。地址的最低位（第一位）可用来区分是单个站地址还是组地址，此低位（第二位）用来区分是全球管理还是本地管理。

* MAC帧格式

前两个字段分别是目的地址和源地址（各6字节），第三个字段类型字段（2字节，表示上一层用的是什么协议），第四个字段数据字段，最后一个字段是4字节的帧检验序列FCS。由于使用曼彻斯特编码，因此，当电压不变时，表示帧结束了。

此外，在帧前面还要插入八个字节的同步字段。（分两部分，前七个字节是同步码，第八字节是帧定界符）。

当MAC帧长数据小于46字节（首尾总共18字节，加上数据部分最少64字节）时，会在数据后面加填充，接收方为了知道填充大小（该丢弃的部分，假如使用IP协议），可用MAC帧的数据长-IP数据报首部信息中的数据长。

* 以太网交换机

提供很多借口，两个主机间通信时，交换机通过交换表查找转发接口。交换表的录入与更新使用自学习算法。

为了增加网路的可靠性，往往会增加一些冗余链路，这就可能使数据兜圈子，未解决这问题，制定了生成树协议STP。

# 网络层

主要内容：

1、虚拟互联网络的概念

2、IP地址和物理地址的关系

3、传统分类的IP地址（包括子网掩码）和无分类域路由间选择CIDR

4、路由选择协议工作原理。

## 1、网际协议IP

与IP协议配套使用的还有三个协议：地址解析协议ARP，网际控制报文协议ICMP，网际组管理协议IGMP。

1. 虚拟互联网络

许多计算机通过路由器相连，路由器只使用下三层协议。网络层使用IP协议。

1. 分类的IP地址

IP地址的编址方法经历了三个阶段：分类的IP地址，子网的划分，构成超网。

四类IP地址

1. IP地址与硬件地址（MAC地址）

MAC地址是物理地址，数据链路层和物理层使用。IP地址是逻辑地址，网络层及网络层以上使用。

1. 地址解析协议ARP

主机ARP高速缓存中存放一个IP地址到MAC地址的映射表，这个表动态更行。具体见P125。ARP是解决同一个局域网上的主机或路由器的IP地址和硬件地址的映射问题。

1. IP数据报格式

看书吧，128页，滑稽。

1. IP层转发分组流程

特定主机路由、默认路由。

当路由器收到一个IP数据报时，查找路由表得知下一跳的地址，然后在数据链路层通过ARP协议获取下一条的MAC地址，把MAC地址填入MAC帧，目的IP地址不变。

## 2、划分子网和构成超网

1. 划分子网

* IP地址的缺点

**IP地址空间的利用率低**有的单位可能申请一个B类地址，但实际接入的不多，为了单位的发展又不愿申请C类地址，这样会造成地址浪费

**给每个物理网络分配网络号会使路由表变得很大**

**两级IP不够灵活**比如某个单位需要在新的地点新开一个网络，可是，在分配到新的IP地址前，新增加的网络是不能工作的。

* 划分子网

为了解决上述问题，IP地址新增一个“子网号字段”，使两级IP地址变成三级。具体划分思路如下：

a、一个拥有许多物理网络的单位，将所属的物理网络划分成多个子网，但这个单位对外表现仍为一个网络。

b、划分子网的方法是从主机号中借用若干位为子网号。

c、其他网络发过来的数据报，先根据网络号找到本单位网络上的路由器，在借由子网号找到子网，最后找到目的主机。

* 子网掩码

就是代表网络号和子网号的位置全1，主机号的位置全0，这货和IP数据报与就能得到网络号和子网号。

**注:**划分子网虽然增加了灵活性，却减少了能连在网络上的主机总数。

* 使用子网时分组间的转发

使用子网划分后，路由表必须有目的网络地址，子网掩码，下一条地址。具体的看书上P140。**（这里有问题，就是那个算法的第四步，为什么当路由表中的目的地址没有时就报错）**

1. 无分类编址CIDR

* CIDR

消除传统的IP地址分类，而将IP地址分为两部分，分别是网络前缀和主机号。CIDR还使用“斜线计法”，在IP地址后面加上斜线“/”，然后写上网络前缀位数。一个单位获得了网络前缀后，还可继续划分子网。

注：个人理解，分类的IP地址网络号必须卡在整数字节处，这样路由表对每个网络号都要记载，而CIDR，路由表只要记住网络前缀就行（不需要卡在整数字节处）。

* 最长前缀匹配

就是当收到的IP数据报匹配多个网络前缀时，选择最长的那个网络前缀作为目的主机所在的网络。

* 二叉线索查找

路由表中的网络前缀用二叉树来存，为了提高二叉线索的查找速度，广泛应用了压缩技术。

## 3、网际控制报文协议ICMP

ICMP是网络层协议。ICMP报文作为IP数据报的数据部分，加上首部后组成IP数据报发送出去。

1. ICMP报文种类

ICMP报文有两种：ICMP差错报文和ICMP询问报文

ICMP前四个字节是固定的，共三个字段：类型（1字节），代码（1字节），检验和（2字节）。之后四字节内容与ICMP报文类型相关，再后面是数据部分，长度取决于ICMP类型。

差错报文：终点不可达；时间超过；参数问题；改变路由

询问报文：会送请求或回答；时间戳请求或回答

1. ICMP应用举例

Ping

## 4、互联网路由选择协议

1. 基本概念

* 理想的路由选择算法具有的特点

**算法必须是正确的与完整的**即沿着各路由表所指引的路由，分组一定能最终到达目的网络和目的主机

**算法在计算上应简单**路由选择的计算不应是网络通信量增加太多的额外开销

**算法能适应通信量和网络拓扑的变化（就是自适应性）**当网络的通信量发生变化时，算法能自适应地改变路由以平衡负载

**算法具有稳定性**在通信量和网络拓扑相对稳定的情况下，算法应收敛于一个可以接受的解，而不应使得路由不停变化

**算法应是公平的**即对所有用户（除了少数优先级高的用户）都是平等的

**算法应是最佳的**“最佳并不是最重要”，主要是让平均时延尽可能小，网络吞吐量尽可能大

* 路由算法分类

根据是否会随网络的通信量或拓扑自适应地进行调整变化来划分，可分为静态路由选择策略和动态路由选择策略。

* 分层次的路由选择协议

现在互联网一般用动态、分布式路由选择协议。互联网把路由选择协议分成两大类，即：

**a、内部网关协议（IGP）**即在一个自治系统内部选择的协议，目前这类路由协议用得最多，如RIP和OSPF

**b、外部网关协议（EGP）**当一个自治系统中的数据报传到另一个自治系统中，当数据报传到一个自治系统边界时，就要使用EGP协议再传送到另一个自治系统。目前最多的是BGP-4协议。

1. 内部网关协议RIP（封装在UDP数据报中）

RIP是一种基于距离向量的路由选择协议。每一个路由器都要维护一个它到目的网络的距离向量，若目的网络是直连的，距离向量就是1，非直连的话，就是其中经过的路由器数+1。RIP协议允许最大经过的路由器数为15，即距离向量最大16，超过的话就是不可达。RIP协议选择的是最短距离路由（到目的网络经过的路由器数量最少），哪怕还存在一条低时延但距离更长的路由。

* RIP协议特点

a、仅和相邻的路由器交换信息

b、路由器交换的信息是当前本路由器知道的全部信息（就是路由表）

c、按固定时间间隔交换信息，当网络拓扑结构发生变化时，路由器也向周围的路由器通告拓扑变化后的路由信息。

路由表的主要信息就是：到某个网络的最短距离以及应进过的下一跳地址。路由表更新的原则就是找到每个目的网络的最小距离。这种算法又叫距离向量算法。

* 距离向量算法

见书P155

* RIP协议报文格式

见书157

（好消息传递快，坏消息传递慢）

1. 内部网关协议OSPF（直接IP数据报）

OSPF最主要的特征是使用分布式的链路状态协议，而不像RIP那样的距离向量协议。

* OSPF基本特点

a、向本自治系统的所有路由器发送信息，这里使用的是洪泛法，就是路由器通过所有的输出端口向所有相邻的路由器发送信息，而每个相邻的路由器又将此信息发送给其所有相邻的路由器（除了刚才发来消息的那个路由器）。

b、发送的信息就是与本路由器相邻的所有路由器的**链路状态**，但这只是路由器知道的部分信息。链路状态就是说明本路由器都和哪些路由器相邻，以及该链路的“度量”。OSPF将这个“度量”用来表示费用、距离、时延、带宽等。

c、只有当链路状态发生变化时，路由器才向所有路由器用洪泛法发送此信息。

由于各路由器之间频繁交换链路状态信息，因此所有路由最终都能建立一个**链路状态数据库**，这个数据库实际上就是**全网的拓扑结构图**。这个拓扑结构图在全网范围内是一致的。每个路由器使用这个数据库中的数据，构建自己的路由表。

OSPF将一个自治系统再划分成为若干个更小的范围，叫做区域。OSPF不用UDP而直接用IP数据报传送。

此外，OSPF还具有以下特点：

a、允许管理员为每条路指派不同的代价

b、如果到同一个目的网路有多种相同代价的路径，那么可以将通信量分配给这几条路径。

c、所有在OSPF路由器之间交换的分组都具有鉴别功能，因而能够保证仅在信赖的路由器之间交换链路状态信息。

d、OSPF支持可变长度的子网划分和无分类编址CIDR

e、由于网络中的链路状态可能经常发生变化，因此OSPF让每一个链路状态都带上一个32位的序号，序号越大状态越新，并且规定链路状态序号的增长速率不得超过每5秒一次。

* OSPF五种类型

类型1：问候分组，用来发现和维持邻站的可达性

类型2：数据库描述分组，向邻站给出自己的链路状态数据库中所有链路状态项目的摘要信息。

类型3：链路状态请求分组，向对方请求发送某些链路状态的详细信息

类型4：链路状态更新分组，用洪泛法对全网更新链路状态。

类型5：链路状态确认分组，对链路更新分组的确认。

OSPF使用的是可靠的洪泛法，即要收到更新分组后发送确认。但是对重复收到的同一个状态分组不进行确认。此外，若N个路由器连接在一个以太网上，每个路由器要向它周围N-1个发送确认，因而共有N(N-1)个链路状态在这个以太网上。为了减少信息量，OSPF采用指定的路由器代表该局域网上所有的链路向连接到该网络上的路由器发送状态信息。

1. 外部网关协议BGP（要建立TCP连接）

在不同的自治系统之间，用BGP而不用内部网关协议的原因是：

第一，互联网规模太大，使得自治系统之间路由选择变得困难。而且，每个自治系统之间的路径度量也可能不一样。

第二、自治系统之间的路由选择必须考虑有关策略。

BGP只能是力求找到一条能到达目的网络且比较好的路由，而并非寻找一条最佳路由。BGP采用路径向量路由选择协议，它与距离向量协议（RIP）和链路状态协议（OSPF）有很大区别。

每个自治系统要选至少一个路由器作为“BGP发言人”，一个BGP发言人与其他自治系统的BGP发言人交换路由信息时，要建立TCP连接。

BGP交换的信息是要到达某个网络（用网络前缀表示）所要经过的一系列自治系统。假设A与B交换路由信息时，B告诉A到达某个目的网络要经过的一条路径（B->……），而A发现自己也在这条路径中，于是就不采取这条路径。

1. 路由器构成

* 路由器结构

路由器是一种具有多个输入端口和多个输出端口的专用计算机，其任务是转发分组。整个路由器的结构可分为两大部分：路由选择部分和分组转发部分。

路由选择部分也叫控制部分，主要构件是路由选择处理机。路由选择处理机的任务就是根据选定的路由选择协议构造出路由表，同时经常或定期的与相邻路由器交换信息维护路由表。

分组转发部分有三部分：交换部分、一组输入端口、一组输出端口。

交换部分是根据交换表把收到的IP数据报从合适的端口转发出去。转发表由路由表得出，转发表必须包括从要到达的目的网络到输出端口和某些MAC地址（如下一跳的MAC地址）的信息。

输入端口有三个模块，分别处理物理层、数据链路层、网络层。若收到的是路由器之间交换信息的分组（如RIP或OSPF分组等），就把他们交给路由选择部分。若收到的是数据部分，就查找转发表从表中找到的输出端口输出。每个输入端口，一般都有一个转发表的副本。路由选择部分负责对这些副本进行更新。

* 交换结构

P170

## 5、IPv6

1. 基本特点

a、**更大的地址空间** IPv6的地址有128位

b、**扩展的地址层次结构** IPv6由于地址空间很大，能划分成更多层次

c、**灵活的首部格式**

d、**改进的选项 ，**IPv6的首部长不变

e、**允许协议进行扩充**

f、**支持即插即用**

g、**支持资源的预分配**

h、**IPv6首部改为八字节对其**

1. 从IPv4到IPv6过渡

下面介绍两种策略：双协议栈和隧道技术

* 双协议栈

是指在完全过度到IPv6前，使一部分主机（或路由器）同时具有IPv4和IPv6的协议栈。它也同时具有两个地址，假设下面有这样的两个路由器A、B，一个IPv4和一个IPv6，它们之间通过IPv4的网络相连，。当一个IPv6的数据报转到A时（目的网络在B那边），A先将IPv6的数据报转到IPv4，当数据报进入B时，B再将IPv4的数据报转成IPv6。但这样会丢失一些首部数据（如流标号）。

* 隧道技术

就是讲IPv6的数据报作为IPv4的数据部分装入IPv4的数据报，这样要在IPv4的协议字段部分设置为41（表示数据部分时IPv6的数据报）。

1. ICMPv6

IPv6也需要通过ICMP来反馈差错信息。新的版本称为ICMPv6，它比ICMPv4要复杂得多，地址解析协议ARP和网际组管理协议IGMP都被合并到ICMPv6中了。

## 6、IP多播

互联网中的多播一般靠路由器来实现，这些路由器必须要加一个能识别多播数据报的软件。能运行多播路由器的称为多播数据报。多播IP数据报的目的地址就是D类地址（IPv4）。多播地址只能用于目的地址，不能用于源地址。

1. 在局域网上进行硬件多播

多播的IP地址是D类地址，而D类地址中能用的位数是28位（D类地址以1110开头）。在硬件地址中，第一个字节的最低位为1时是多播地址。IANA拥有的以太网地址的高24位是确定的（可用的有24位？），其中多播地址数占一半。

多播硬件地址只有23位可用，多播硬件地址的23位和IP地址的23位对应，二IP地址28位可用，因此，IP地址的前5位不能用来构成以太网硬件地址。

1. 网际组管理协议IGMP与多播路由选择协议

* IP多播需要两种协议

IGMP不知道IP多播组的成员数，也不知道这些成员分布在哪些网络上。IGMP是让连接在本地局域网上的多播路由器知道本局域网上是否有主机（严格讲，是主机的某个进程）参加或退出某个多播组。

仅有IGMP协议是不能完成多播任务的。连接局域网上的路由器还需要与互联网上的其他多播路由器协同工作，以便把多播数据报用最小的代价发送给多播组所有成员，这需要用到**多播路由选择协议。**

单播路由选择协议通常是在网络拓扑发生变化时才需要更新路由，而多播路由选择协议则不是（假设已联网的某台主机加入到某个多播组，这时候网络拓扑结构未变化）。

多播数据报可由未加入多播组的主机发出，也可通过没有组成员接入的网络。

* 网际组管理协议IGMP

IGMP用IP数据报传递报文（即IGMP报文加上IP首部构成IP数据报）。IGMP是IP协议的组成部分。

IGMP工作分两阶段：

a、当某主机加入新的多播组时，该主机向多播组的多播地址发送一个IGMP报文，声明自己成为该组的成员。本地的多播路由器收到IGMP报文后，还要利用多播路由选择协议把这种组成员关系转发给互联网上的其他多播播放器。

b、组成员关系是动态的。本地多播路由器要周期探询本地局域网上的主机，以便知道这些主机是否还继续是组的活跃成员。只要有一台主机对某个组响应，那么多播路由器就认为这个组是活跃的。若一个组在经过几次探询后没有一台主机响应，多播路由器就认为本网络上的主机都离开了这个组，因此也就不再把这个组的成员关系转发到其他组。

* 多播路由选择协议

多播路由实际上就是要找出以源主机为根节点的多播转发树。在多播转发树上，每一个多播路由器向树的叶节点转发收到的多播数据包，但多播转发树上的路由器不会收到重复的多播数据报。

已有多种多播路由选择协议，它们转发多播数据报时一般用以下方法：

**a、洪范和剪除**

此方法适用于较小的多播组。路由器转发多播数据报是采用洪范的方法，为了不重复转发一个多播数据报，采用了一种叫反向路径广播RPB的策略，该策略思想如下：每个路由器在收到一个多播数据报时，检查该数据报是不是通过最短路劲发过来的，检查的方法就是看发送的数据报的路由器在源点到本路由最短路径上，是不是本路由器的上一个路由器。若是通过最短路径发来的，就接收，如不是，就丢弃。若存在多条最短路径，则选择IP最小的那一个。

如果某个多播路由器发现它的下游树枝已经没有该多播组成员，就把它和它的下游树枝一起剪除。

1. **隧道技术**

此方法适用于地理位置很分散的多播组。若多播路由器R1和R2之间的网络不支持多播，那么他们之间的多播数据报就要加上一个首部封装成一个普通的IP数据报（单播），来经过这个网络。

**基于核心的发现技术**

这种方法对多播组的大小在较大范围内变化都适合。首先一个多播组G指定一个核心路由器。当有一个路由器R1

要向这个核心路由器发送数据报时，途经的每个路由器都要检查其内容，若数据报到达路由器R2，R2按一下方式处理。若数据报是多播数据报，且目的地址是G的组地址，R2就像G的所有成员转发；若是请求加入多播组G的数据报，R2以后就向R1用隧道技术转发每一个G的多播数据报副本。

# 运输层

主要内容：

1、运输层为相互通信的应用进程提供逻辑通信

2、端口与套接字的特点

3、无连接UDP的特点

4、面向连接的TCP特点

5、在不可靠的网络上实现可靠传输的工作原理，停止等待协议和ARP协议。

6、TCP的滑动窗口、流量控制、拥塞控制和连接管理。

## 1、运输层协议概述

从运输层的角度看

通信的端点不是主机，而是主机的进程。

运输层复用与分用。

运输层对收到的报文进行差错检测，网络层只进行首部的差错检测。

1. UDP

* 特点

无连接，不可靠交付，面向报文，无拥塞控制，支持一对一、一对多、多对一，UDP首部开销小。

关于面向报文：即应用层交给UDP多长的报文，UDP都照样发送，一次发送一个报文。因此，应用层必须选择合适大小的报文，若报文太长，到了网络层会被分片，若太短，则到网络层后，网络层首部太大。

* 首部格式

UDP由数据部分和首部部分组成，首部有四个字段，每个字段两个字节长。各字段意义如下：

a、**源端口** 源端口号，在对方需要回信的时候用的上。

b、**目的端口**  目的端口号，在终点交付报文时用得上。

c、**长度** UDP用户数据报长度，最小值是8（仅有首部）

d、**检验和** 12字节伪首部，8字节UDP首部，以及数据部分参与计算。

1. TCP

* 特点

面向连接，每个TCP连接只能有两个端点，可靠交付，全双工通信，面向字节流。

对于面向字节流：虽然应用程序和TCP交互式一次一个数据块（大小不等），但TCP把交下来的数据仅仅看成一连串无结构的字节流。

* TCP连接

TCP连接的端点不是主机，不是主机IP，不是应用进程，也不是运输层的协议端口。TCP连接的端口交套接字（socket）或者插口。端口号拼接在IP地址后就构成了套接字。每一条TCP连接唯一被两个端点（套接字）确定。

## 2、可靠传输的原理

理想的传输协议有两个特点：传输信道不产生差错；不管发送方以多快的速度发送数据，接收方总能来得及处理接收到的数据。

1. 停止等待协议（ARQ）

假设A是发送方，B是接收方。

* B对A发送的每个报文，都要发报文对A进行确认。A在收到确认报文后发送下一个报文。
* 若某报文在发送过程中出现差错，B收到后就丢弃，并什么也不做。当A在一定时间没收到确认报文后，就对上一个发送的报文再次发送。这里有三个点要注意：

a、A发送完一个报文后，要暂时保存该报文的副本，因为可能重传

b、报文和确认报文都要进行编号

c、重传时间应当比数据在分组传输的平均往返时间更长

* 若B发向A的确认丢失了时，A也需要向B再次发送上个数据报，B接收到后，就丢弃该数据报，但仍要发送确认。A若是收到了B对该数据报第一次发送的确认（迟到的），就什么也不做。

1. 连续ARQ协议

发送方维持一个滑动窗口。

接收方采用累积确认的方式，即对按序到达的最后一个分组发送确认。

## 3、TCP报文的格式

TCP报文首部前20个字节固定，后面4n个字节是需要增加的选项。

a、**源端口和目的端口** 各占两2个字节。

b、**序号** 4字节，整个要传送数据的第一个字节号

c、**确认号** 4字节，期望收到的下一个报文段的第一个数据字节的序号

d、**数据偏移** 4位，数据部分起始处距TCP报文起始处的距离，单位是4字节

e、**保留** 6位，保留以后使用

f、**紧急URG**  1位，该位为1时紧急指针字段有效。

g、**确认ACK** 1位，该位为1，确认号字段有效

懒得写了，翻书吧，P219

## 4、TCP可靠传输的实现

发送方维持一个发送窗口，接收方维持一个接收窗口，窗口里的数据以字节为单位。发送窗口里的是已经发送但还未收到确认的字节以及还未发送的字节。接收窗口里的是还未收到的字节以及未按序收到的字节。

接收方需要对发送方发出确认报文，里面的内容大致是希望收到的下一个字节的序号（表明这之前的都收到了），发送方收到确认后，按照确认报文里的序号来移动窗口（即将已经确认收到了的报文移出发送窗口）。

## 5、流量控制（拥塞控制）

TCP进行拥塞控制的算法有四种：慢开始、拥塞避免、快重传和快恢复。

1. 慢开始和和拥塞避免

发送方维持一个叫做拥塞窗口cwnd的状态变量。拥塞窗口的大小取决于网络的拥塞程度。发送方让自己的发送窗口等于拥塞窗口。

发送方由小到大增加发送窗口，起初， ·先将初始拥塞窗口设置为2至4个发送方的最大报文段SMSS的值，每收到一个新的报文段的确认后，可以把拥塞窗口增加最多一个SMSS的数值。（TCP以字节数作为窗口大小的单位）

在**慢开始**时，没收到一个报文段的确认，拥塞窗口就+1，这样一个轮次过后，拥塞窗口cwnd就加倍了。

为了防止拥塞窗口cwnd增长过快而引起网络堵塞，还需要设置一个慢开始门限（ssthresh）。

当cwnd<ssthresh，采用慢开始算法；

当cwnd>ssthresh，采用拥塞避免算法；

当cwnd=ssthresh，两者都可以

**拥塞避免算法**是每经过一个轮次，拥塞窗口只+1，而不是成倍增长。

当网络出现超时时，发送方判断为网络拥塞，之后，门限值ssthresh减半，同时拥塞窗口cwnd=1，进入慢开始阶段。

当接收方收到失序的报文段时，也要立即发送对已收到报文的确认，当发送方收到三个重复的确认报文时，就知道那个报文丢失了，于是立即重传（“快重传”），之后，不进行慢开始算法，而是进行快恢复算法，门限值等于当前窗口值得一半，然后窗口值等于门限值，并开始执行拥塞避免算法。

补：其实，接收方也维护着一个接收窗口rwnd，并把这个值写进TCP首部，来通知发送方，发送方的窗口不能超过这个值。

1. 主动队列管理AQM

路由器对处理IP数据报时采用的是先进先出队列，当队列已经满时，以后到达的分组都被丢弃。这样一来发送方收不到确认，从而进入慢开始。而且，在网络层中有很多TCP连接，它们都复用在IP数据报里，这种情况下，若分组被丢弃，就会使很多TCP连接进入慢开始（全局同步），之后又会突然增大。

为了避免全局同步，提出了主动对列管理（AQM），AQM有几种不同的实现方法，下面介绍随机早期检测（RED）。

RED要求路由器维护两个参数，即对列长度的最小门限a和最大门限b。当一个数据报到达时，计算平均队列长度x:

若x<a，则放进队列

若x>b。则丢弃

其他，按某一概率丢弃该数据报。

## 6、TCP连接管理

1. TCP连接建立（三次握手）

假设A时客户机（主动建立连接），B是服务器，被动建立连接。

建立连接时，A先发送请求报文，这时同步位SYN=1，同时选择一个初始序号seq=x。TCP规定，SYN=1的报文段（同步报文）不能携带数据，但要消耗掉一个序号。

B收到请求后，若同意建立连接，就向A发送确认，确认报文中，确认报文中SYM与Ack字段都为1，确认号ack是x+1，同时要为自己选一个初始序号seq=y。这个报文不能携带数据，但要消耗一个序号。

A收到B的确认后，再给B发送确认。ACK字段为1，确认号ack=y+1，自己的序号是x+1。这个报文可以带数据，但如果不带数据，就不消耗序号。这之后，TCP连接就建立了。

上述就是三次握手。B发送给A的报文段也可以拆成两个，先发送一个确认报文段（ACK1,ack=x+1），再发送一个同步报文段（SYN=1，seq=y），这样就变成了四报文握手。

1. TCP连接的释放

通信双方都可释放连接。现在假设A的应用进程先发送连接释放报文段。（如B先发送，则角色相反）

A的连接释放报文段的FIN位为1，序号seq=u（u是上一个传送的字节序号+1）。这时A进入**FIN-WAIT-1**的阶段。注意，及时FIN报文段不带数据，也要消耗一个序号。

B收到连接释放报文段后发出确认，确认号ack=u+1，假设这个报文段序号seq=v（B传送的数据的前一个字节序号+1），然后B就进入了**CLOSE-WAIT。**这个时候处于半关闭的状态，即A没有数据发给B，但是B要发数据的话A还是要接收。

A收到B的确认报文后，就进入**FIN-WAIT-2**的状态，等待B发送释放报文段。

若B没了向A发送的数据，B就发送连接释放报文，该报文FIN=1，序号为w，确认号为u+1。之后B进入**LAST-ACK**阶段。

A收到B的释放报文后，对其进行确认，报文中ACK=1，ack=w+1],seq=u+1。然后进入**TIME-WAIT**状态，此时TCP连接还没释放，再经两个最长报文寿命（MSL）后连接释放。

**补：**TIME-WAIT的作用

1. 为了保证A发送的最后一个ACK报文能到达B。若B没收到这个报文，就会重传FIN+ACK报文。
2. 防止“已失效的连接请求报文段”出现在连接中。A发送最后一个报文段后，等待2\*MSL的时间，确保本次连接的所有报文都消失在网络。

# 应用层

应用层协议应当定义：

1. 应用进程报文的类型，如请求报文和响应报文
2. 各种报文类型的语法，如报文各个字段的详述。
3. 字段的语义
4. 进程何时、如何发送报文，以及如何响应报文。

本章主要内容：

1、域名DNS

2、万维网和HTTP协议，以及万维网两种不同的搜索引擎

3、电子邮件传送过程，SMTP协议和POP3协议、IMAP协议使用的场合。

4、动态主机配置协议DHCP

5、网络管理的三个组成部分（SMTP，SMI，MIB）

6、系统调用和应用编程接口的基本概念

7、P2P文件系统

## 1、域名系统

1. 概述

域名系统是将域名解析为IP地址的系统。互联网上的域名系统被设计成一个联机分布式的数据库，DNS大多名字都在本地进行解析。

域名到IP地址解析：当某一应用进程要将域名解析为IP地址时，就调用解析程序，并成为DNS的一个客户，将带解析的域名放入DNS请求报文，以UDP数据报的形式发送到本地域名服务器。本地域名服务器查找到IP地址后，将IP地址放入回答报文中返回。若本地域名服务器不能回答该请求，就要向上级域名服务器查询。

1. 域名服务器

一个DNS服务器管辖一个区。每一个区设置相应的权限域名服务器，从来保存该区所有域名到IP地址的映射。区是域的子集（区能小于等于域，但不能大于域）。

域名服务器的类型：

1. 根域名服务器 最重要的域名服务器，所有根域名服务器都知道所有顶级域名服务器的域名和IP地址。当互联网上任何一个本地域名服务器无法解析某个域名时，首先就要求助根域名服务器。在许多情况下，根域名服务器并不直接告诉IP地址，而是告诉下一步要去那个DNS服务器查找。
2. 顶级域名服务器 负责管理在该顶级域名服务器下注册的二级域名
3. 权限域名服务器 就是前面讲过的一个区的域名服务器。当权限域名服务器不能给出最后应答时，就会告诉发出请求的DNS客户，下一步去哪查。
4. 本地域名服务器 当一个主机发送DNS请求时，先送到的就是本地域名服务器。
5. 域名解析过程

* 主机向本地域名服务器的查询一般是递归查询，即：主机所询问的本地域名服务器不知道IP地址，那本地域名服务器就以DNS客户的身份向其他根域名服务器进行查询，而不是让主机查询。
* 本地域名服务器向根域名服务器查询一般是迭代查询。若根域名知道IP，就告诉IP，不知道的话就告诉下一步该往哪个顶级域名服务器查询；若顶级域名服务器知道IP，就告诉IP，若不知道，就告诉下一步查询的域名服务器。

补：本地域名服务器也可采用递归查询，这取决于最初的查询请求报文的设置。

为了提高查询效率，在域名服务器中广泛地采用高速缓存。

## 2、文本传输协议

1. 文件传输协议FTP概述

FTP是基于TCP的，TFTP是基于UDP的，它们都是文件共享协议的一大类，即复制整个文件，有个共同的特点：先取得一个本地的文件副本，之后的修改就是在本地文件副本上进行修改，然后将修改后的文件副本传回原节点。

1. FTP工作原理

一个FTP进程可为多个客户进程提供服务。FTP服务器进程有两部分组成：一个主进程，负责接收新的请求；另外若干个从属进程，负责处理单个请求。

主进程工作步骤：

1. 打开熟知端口21
2. 等待客户进程发出连接请求
3. 启动从属进程处理连接请求。从属进程在处理过程中可创建其他子进程，从属进程处理完毕后终止。
4. 回到等待状态

当客户端进程发出连接请求后，服务端开启两个从属进程：控制进程和数据传送进程。控制进程用来传输连接的控制信息，数据传送进程传送文件。

FTP在传送大文件时，由于要复制整个文件到本地，因此效率不是很好。然而，**网络文件系统NFS**采用另一种思路，即只拷贝很小的一个片段，而不复制整个文件。

1. 简单文件传送协议TFTP

TFTP支持文件传输而不支持交互。主要优点有两个：1、支持UDP传输，当需要将程序或文件同时向许多机器下载时需要TFTP。2、TFTP占用内存小。

TFTP主要特点：

1. 每次传送的报文中有512字节数据，最后一次可不足512字节
2. 数据报文按序号编号，从1开始
3. 支持ASCII码或二进制传送
4. 可对文件进行读写
5. 使用很简单的首部

TFTP很像停止等待协议，发送完一个文件块就要求确认。TFTP传送的文件若刚好是512字节的整数倍，则在最后还要发一个只含头部不含数据的报文；若不是整数倍，则在最后发送文件块少于512字节的报文。

## 3、万维网WWW

1. WWW概述

万维网必须解决的问题：

1. 如何标志分布在互联网上的万维网文档
2. 用怎样的协议来实现万维网上的各种链接
3. 怎样使不同作者创作的不同风格的万维网文档都能在互联网上的主机显示出来。
4. 怎么使用户能很方便查找信息

对于1，使用统一的资源定位符URL，对于2，使用超文本传输协议HTTP，对于3，使用HTML语言。

1. URL

<协议>://<主机>:<端口>/<路径>

1. HTTP

* 操作过程

HTTP协议本身无连接，虽然使用了TCP连接，但通信双方在交换HTTP报文前不需要先建立HTTP连接。

HTTP是无状态的，同一个客户第二次访问服务器页面时，服务器的响应与第一次相同。

当用户点击某个万维网文档时，HTTP协议先和服务器建立TCP连接，在第三次握手时，将HTTP请求报文作为第三次握手的数据部分。服务器收到请求后，将请求的文档作为响应报文返回给用户。

HTTP1.0是**非持续连接**的，即每次请求都要建立一个新的TCP连接，HTTP1.1是**持续连接**。

* 代理服务器

代理服务器将最近的一些请求和响应暂存到本地磁盘。以提高响应速度。