

XI'AN JIAOTONG UNIVERSITY

# 数据通信与计算机网络

## 第6章 网络互连

XI'AN JIAOTONG UNIVERSITY

### 主要内容与基本要求

**主要内容**

- 互连网的概念、中间设备及路由器的作用；
- 网际协议IP，IP地址的演进过程，地址转换；
- IP数据报格式、分片操作，IP层处理数据报的流程，ICMP；
- 因特网路由选择协议，自治系统的概念，两类路由选择协议；
- 多播IGMP、虚拟专用网及网络地址转换，IPv6。

**基本要求**

- 理解互连网的概念，路由器的作用；
- 掌握网际协议IP，分类的IP、划分子网、构造超网、地址转换；
- 理解IP数据报的格式，掌握IP层处理数据报的流程、分片方法；
- 掌握AS的概念、内部网关协议RIP和OSPF，了解外部网关协议；
- 理解ICMP；了解IGMP、VPN、NAT、IPv6和ICMPv6。

2011-8-31 《数据通信与计算机网络》——网络互连 2

XI'AN JIAOTONG UNIVERSITY

### 本章目录

- 6.1 路由器在网际互连中的作用 ▶
- 6.2 因特网的网际协议IP ▶
- 6.3 划分子网和构造超网 ▶
- 6.4 因特网控制报文协议ICMP ▶
- 6.5 因特网的路由选择协议 ▶
- 6.6 IP多播和因特网组管理协议IGMP ▶
- 6.7 虚拟专用网VPN和网络地址转换NAT ▶
- 6.8 下一代的网际协议IPv6(IPng) ▶

2011-8-31 《数据通信与计算机网络》——网络互连 3

XI'AN JIAOTONG UNIVERSITY

### 6.1 路由器在网际互连中的作用

路由器是把多种不同类型的网络互相连接在一起形成因特网的关键设备。其主要作用是在不同类型的网络之间转发分组。

#### 6.1.1 路由器的构成

直接交付和间接交付的概念。

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。路由器工作在体系结构的第三层。

**路由器的典型结构**

整个路由器结构可以分为路由选择和分组转发两大部分。路由选择部分也叫控制部分，其核心是路由选择处理机。分组转发部分又可分三个部分：交换结构、一组输入端口和一组输出端口。

2011-8-31 《数据通信与计算机网络》——网络互连 4

XI'AN JIAOTONG UNIVERSITY

### 6.1 路由器在网际互连中的作用

交换结构（switching fabric）的作用就是根据转发表把从某个端口进入的分组转发到一个合适的输出端口。

“转发”与“路由选择”的区别。路由表与转发表的关系：路由表是由路由算法计算得到的，一般仅包含从目的网络到下一跳的映射；而转发表是从路由表得出的，它必须包含完成转发功能所必需的信息，其每一行必须包含从目的网络到输出端口和某些硬件地址信息的映射。

输入端口中的查找和转发功能在路由器的交换功能中是最重要的。为了使交换功能分散化，往往将复制的转发表放在每个输入端口中。

查找转发的概念很简单，具体实现的困难在于如何实现很高的转发速率。（百万分组每秒，Mpps）

路由器输入、输出端口对分组的处理过程如下：

2011-8-31 《数据通信与计算机网络》——网络互连 5

XI'AN JIAOTONG UNIVERSITY

### 6.1 路由器在网际互连中的作用

从线路接收分组

输入端口的处理

物理层处理

数据链路层处理

网络层处理  
分组排队  
查表和转发

交换结构

输出端口的处理

网络层处理  
分组排队  
缓存管理

数据链路层处理

物理层处理

向线路发送分组

2011-8-31 《数据通信与计算机网络》——网络互连 6

### 6.1 路由器在网际互连中的作用

#### 6.1.2 交换结构

交换结构是路由器的关键构件。下面介绍三种常用的实现方法。

- (1) 基于存储器的交换结构，交换速率小于存储器带宽的一半；
- (2) 基于总线的交换结构，交换速率小于总线速率；

基于存储器的交换结构                      基于总线的交换结构

2011-8-31                      《数据通信与计算机网络》——网络互连                      7

### 6.1 路由器在网际互连中的作用

(3) 基于互连网络的纵横交换结构，比基于总线的结构有更高的交换速率。

基于互连网络的交换结构

2011-8-31                      《数据通信与计算机网络》——网络互连                      8

### 6.1 路由器在网际互连中的作用

#### 6.1.3 互联网与因特网

互连在一起的网路要进行通信，会遇到很多问题，如：不同的寻址方案、不同的网络接入机制、不同的差错恢复方法、不同的状态报告方法、不同的路由选择技术、不同的服务及不同的管理与控制方式等。

把网络互相连接起来需要一些中间设备，又称中间系统或中继系统。

- (1) 物理层中继系统，转发器 (repeater)；
- (2) 数据链路层中继系统，网桥或桥接器 (bridge)；
- (3) 网络层中继系统，路由器 (router)；
- (4) 网桥和路由器的混合物，桥路由器 (brouter)；
- (5) 网络层以上的中继系统称为网关 (gateway)。

注意：早期曾把路由器称为网关。名词“Internet”与“internet”。

2011-8-31                      《数据通信与计算机网络》——网络互连                      9

### 6.2 因特网的网际协议IP

IP是TCP/IP协议体系中两个重要的协议之一。与其配套使用的还有四个协议。地址解析协议ARP、逆地址解析协议RARP、因特网报文控制协议ICMP及因特网组管理协议IGMP。

应用层	各种应用层协议 (TELNET, FTP, SMTP 等)
运输层	TCP, UDP
网际层	ICMP, IGMP IP RARP, ARP
网络接口层	与各种网络接口
	物理硬件

2011-8-31                      《数据通信与计算机网络》——网络互连                      10

### 6.2 因特网的网际协议IP

#### 6.2.1 分类的IP地址

1. IP地址及其表示方法

IP地址就是给每个连接在因特网上的主机（或路由器）分配的在全世界范围惟一的32bit标识符。由因特网名字与号码指派公司ICANN分配。

IP地址的编址经历了三个阶段：分类的IP地址、划分子网和构造超网。

分类的IP地址将IP地址划分为若干个固定类，每类都由两部分组成：

IP地址 ::= {<网络号>, <主机号>}                      (6-1)

**分类的IP地址**

IP地址指出的是连接到某个网络上的某个主机。

分类的IP地址出发点是适应大小不同的网络。

转发表只使用IP地址中的网络号 (net\_id) 来查找路由。

2011-8-31                      《数据通信与计算机网络》——网络互连                      11

### 6.2 因特网的网际协议IP

使用分类的IP地址时，路由器转发分组的步骤：

- (1) 先按IP地址中的net\_id找到目的网络；
- (2) 当分组到达目的网络后，用主机号host\_id把数据报直接交付给主机。

IP地址常用点分十进制记法 (dotted decimal notation)

1000000000001011000001100011111    # 32bit二进制

10000000 00001011 00000011 00011111    # 插入空格提高可读性

128      11          3          31          # 转换为十进制

使用点分十进制记为：128.11.3.31。

2. 常用的三种类别的IP地址

几个特殊的规定：net\_id全“0”表示这个，即本网络。

2011-8-31                      《数据通信与计算机网络》——网络互连                      12

### 6.2 因特网的网际协议IP

host\_id全“0”表示网络地址；全“1”表示所有的，用于网内广播。  
net\_id为127的A类地址保留用于回环测试（loopback test）。  
所有，各类IP地址可以使用的范围如下表所示。

网络类别	最大网络数	第一个可用的网络号	最后一个可用的网络号	每个网络中的最大主机数
A	126 ( $2^7-2$ )	1	126	16 777 214
B	16384 ( $2^{14}$ )	128.0	191.255	65 534
C	2 097 152 ( $2^{21}$ )	192.0.0	223.255.255	254

IP地址的重要特点：  
(1) IP地址是一种分层次的地址结构。只分配网络号；路由器只根据网络号转发分组，减少了路由表所占的存储空间。

2011-8-31 《数据通信与计算机网络》——网络互连 13

### 6.2 因特网的网际协议IP

(2) IP地址实际上是标志一个主机（或路由器）和一个链路的接口。  
(3) 用转发器或网桥连接起来的若干个局域网仍为一个网络。  
(4) 所有分配到网络号net\_id的网络都是平等的。

#### 6.2.2 IP地址与硬件地址

2011-8-31 《数据通信与计算机网络》——网络互连 14

### 6.2 因特网的网际协议IP

IP地址放在IP数据报的首部，而硬件地址放在MAC帧的首部。在网络层和网络层以上使用的是IP地址，而数据链路层及以下使用的是硬件地址。数据链路层看不见数据报的IP地址。

#### IP地址和MAC地址

问题：主机或路由器怎样知道应当在MAC帧的首部填入什么样的硬件地址？路由器的路由表是怎样得出的？

#### 6.2.3 地址解析协议ARP和逆地址解析协议RARP

IP地址不能直接用来进行通信。不管网络层使用的是什么样的协议，在实际网络的链路上传送数据帧时，最终还是必须使用硬件地址。  
每个主机都设有一个ARP高速缓存（ARP cache），其中存放的是本主机目前知道的局域网各主机或路由器的IP地址到硬件地址的映射表。

2011-8-31 《数据通信与计算机网络》——网络互连 15

### 6.2 因特网的网际协议IP

若高速缓存中查不到硬件地址，则需要通过ARP请求分组。

#### ARP请求分组和响应分组

ARP缓存中的每个映射项目都设有生存时间。  
注意：ARP解决的是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。  
地址解析的过程是自动进行的。  
那么，为什么不直接使用硬件地址通信呢？  
不同的网络使用不同的硬件地址，若直接使用硬件地址通信，就必须进行非常复杂的硬件地址转换工作，这几乎是不可能的。  
统一的IP地址解决了这个问题。  
逆地址解析协议RARP使只知道自己硬件地址的主机能知道其IP地址。

2011-8-31 《数据通信与计算机网络》——网络互连 16

### 6.2 因特网的网际协议IP

#### 6.2.4 IP数据报的格式

IP数据报的格式能清楚地说明IP协议都具有什么功能。一个IP数据报由首部和数据两部分组成，首部又分为固定长度部分和可变长度部分。

##### 1. IP数据报首部的固定部分中的各字段

#### IP数据报首部固定部分中各字段

常用协议的协议字段值如下：

协议名	ICMP	IGMP	TCP	EGP	IGP	UDP	IPv6	OSPF
协议字段值	1	2	6	8	9	17	41	89

2011-8-31 《数据通信与计算机网络》——网络互连 17

### 6.2 因特网的网际协议IP

#### 2. IP数据报首部的可变部分

IP首部的可变部分就是一个用来支持排错、测量以及安全等措施的选项字段。可变部分的长度从1到40字节不等，但应为4字节的整数倍。  
可变部分是为了增加数据报的功能，但首部长度可变增加了路由器处理数据报的开销。实际上很少使用这些选项。

#### 6.2.5 IP层转发分组的流程

路由器和结点交换机非常相似，两者的区别如下：  
(1) 路由器用来连接不同的网络，交换机在一个特定的网络中工作。  
(2) 路由器专门转发分组，交换机可以连接许多主机。  
(3) 路由器使用IP协议，交换机使用所在广域网的特定协议。  
(4) 路由器根据目的网络找下一跳（即下一个路由器），交换机根据目的站所接入的交换机号找下一跳（即下一个交换机）。

2011-8-31 《数据通信与计算机网络》——网络互连 18

### 6.2 因特网的网际协议IP

**一个简单的路由表**

我们根据目的网络地址确定下一跳路由器IP。这样做的结果是路由器首先设法找到目的主机所在目的网络上的路由器，当到达最后一个路由器时，才向目的主机直接交付。

**特定主机路由**，允许对特定的目的主机指明一个路由。目的是可以使网络管理人员能更方便地控制网络和测试网络。当然，也可以用于某些需要考虑安全的场合。

路由器也可采用默认路由减少路由表所占空间及搜索时间。

**注意：**路由器从路由表中得到下一跳路由器的IP地址后，不是将其填入IP数据报，而是送交网络接口软件，由网络接口软件负责用ARP把IP地址转换成硬件地址，并将硬件地址放在MAC帧的首部。

2011-8-31 《数据通信与计算机网络》——网络互连 19

### 6.2 因特网的网际协议IP

因特网中某个路由器的IP层执行的分组转发算法如下：

- (1) 从数据报首部提取目的IP地址D，得到目的网络地址N。
- (2) 若N是与本路由器直接相连的某个网络，直接交付；否则执行(3)。
- (3) 若路由表中有目的地址D的特定主机路由，则把数据报发给指定的下一跳路由器；否则执行(4)。
- (4) 若路由表中有到达网络N的路由，则把数据报转交给路由表中指明的下一跳路由器；否则执行(5)。
- (5) 若路由表中有个默认路由，则将数据报发送给默认路由器；否则执行(6)。
- (6) 报告转发分组出错。

**Example...**

2011-8-31 《数据通信与计算机网络》——网络互连 20

### 6.3 划分子网和构造超网

#### 6.3.1 划分子网

1. 从两级IP地址到三级IP地址

分类的IP地址设计不够合理，主要有：

- (1) IP地址空间的利用率有时很低。
- (2) 给每个物理网络分配一个网络号会使路由表变得太大，从而使网络性能变坏。
- (3) 两级的IP地址不够灵活。

为解决这些问题，提出了划分子网（subnetting）的概念。其基本思路是：

(1) 划分子网是一个单位内部的事情，该单位对外仍然表现为一个没有划分子网的网络。

2011-8-31 《数据通信与计算机网络》——网络互连 21

### 6.3 划分子网和构造超网

(2) 划分子网的方法是从主机号字段借用若干比特作为子网号字段 subnet\_id。所以两级的IP地址在本单位内部就变为三级的IP地址，记为：

$$\text{IP地址} ::= \{ \langle \text{网络号} \rangle, \langle \text{子网号} \rangle, \langle \text{主机号} \rangle \} \quad (6-2)$$

(3) 从其它网络发给本单位某主机的IP数据报，仍然根据数据报中的目的地址的net\_id字段找到连接在本单位网络上的路由器。该路由器在收到IP数据报后，按net\_id和subnet\_id找到目的子网后将IP数据报交付给目的主机。

**划分子网的例子**

划分子网只是把IP地址的本地部分（即主机号字段host\_id）进行再划分，而不改变IP地址的因特网部分（即网络号字段net\_id）。

2011-8-31 《数据通信与计算机网络》——网络互连 22

### 6.3 划分子网和构造超网

#### 2. 子网掩码

子网上的主机能否通过本子网上的路由器与外界直接通信呢？答案是肯定的，但需要引入子网掩码（subnet mask）的概念。

	因特网部分		本地部分	
两级 IP 地址	网络号 net-id		主机号 host-id	
三级 IP 地址	net-id		subnet-id	host-id
子网掩码	1111111111111111		11111111	00000000
划分子网时的网络地址	net-id		subnet-id	host-id 为全 0

2011-8-31 《数据通信与计算机网络》——网络互连 23

### 6.3 划分子网和构造超网

在划分子网的情况下，网络地址（即子网地址）就是将主机号 host\_id 置为0的IP地址。即将子网掩码和IP地址逐比特相“与”的结果。

使用子网掩码后，不管网络是否划分子网，只要将子网掩码和IP地址进行逐比特的“与”运算，就可以得到网络地址。划分子网后的路由器就是这样处理分组的。

一个问题：在不划分子网时，为什么还要用子网掩码？

对不划分子网的A、B、C类网络使用默认子网掩码。

子网掩码是一个网络或一个子网的重要属性。路由表中的每个项目，除了要给出目的网络地址外，必须同时给出该网络的子网掩码。

书上195页表6-6。子网号字段的长度固定。

关于全“0”和全“1”的子网号的使用问题。

划分子网增加了灵活性，但却减少了能连接在网络上的主机数。

2011-8-31 《数据通信与计算机网络》——网络互连 24

**6.3 划分子网和构造超网**

**6.3.2 使用子网掩码的分组转发过程**

使用子网掩码后，路由器的分组转发算法也要做出相应的改动。

如使用两级IP地址时分组转发算法的第(1)条是：“从数据报的首部提取目的站IP地址D，得出目的网络地址为N”。对两级IP地址，只要判断首字节的前1~3位可以很容易地得到网络地址。

但在划分子网的情况下，从IP地址不能得出唯一的网络地址来，因为仅从IP地址中我们无法知道子网号字段的长度，必须在子网掩码的帮助下才能得到这个信息，而数据报的首部没有提供子网掩码信息。

**使用子网掩码的分组转发过程**

所以，划分子网时路由器转发分组的算法是：

2011-8-31 《数据通信与计算机网络》——网络互连 25

**6.3 划分子网和构造超网**

- (1) 从收到的分组的首部提取目的IP地址D。
- (2) 先用与本路由器直接相连的各网络的子网掩码和D逐比特相“与”，看结果是否和相应的网络地址匹配。若匹配，则将分组直接交付。否则就是间接交付，执行(3)。
- (3) 若路由表中有目的地址为D的特定主机路由，则将分组传送给指明的一跳路由器；否则，执行(4)。
- (4) 对路由表中的每一行，将其中的子网掩码和D逐比特相“与”，若其结果与该行的目的网络地址匹配，则将分组传送给该行指明的一跳路由器；否则，执行(5)。
- (5) 若路由表中有一个默认路由，则将分组传送给路由表中所指明的默认路由器；否则，执行(6)。
- (6) 报告转发分组出错。

2011-8-31 《数据通信与计算机网络》——网络互连 26

**6.3 划分子网和构造超网**

**6.3.3 无分类编址CIDR**

1. 网络前缀

划分子网在一定程度上缓解了因特网发展中遇到的困难。然而1992年，因特网面临三个必须尽早解决的问题：

- (1) B类地址将要全部分配完毕；
- (2) 因特网主干网上的路由表中的项目急剧增长；
- (3) IPv4的地址空间最终将耗尽。

早在1987年，RFC 1009就指明在一个划分子网的网络中可同时使用几个不同的子网掩码，即变长子网掩码VLSM (Variable Length Subnet Mask)，以进一步提高IP地址资源的利用率。在VLSM的基础上，提出了无分类域间路由选择CIDR (Classless Inter-Domain Routing) 编址方法。

CIDR的主要特点有两个：

2011-8-31 《数据通信与计算机网络》——网络互连 27

**6.3 划分子网和构造超网**

- (1) 消除了传统的A类、B类和C类地址以及划分子网的概念，从而可以更加有效地分配IPv4的地址空间。

CIDR使用各种长度的“网络前缀” (network-prefix) 代替分类地址中的网络号和子网号，使IP地址从三级编址又回到了两级编址，但变成了无分类的两级编址。记为：

IP地址 ::= {<网络前缀>, <主机号>} (6-3)

- (2) CIDR把网络地址前缀都相同的连续的IP地址组成“CIDR地址块”。一个地址块由起始地址和地址块中的地址数来定义。

CIDR的“斜线记法” (slash notation)，如128.14.46.34/20。主机号字段的全0和全1不使用。

CIDR虽然抛弃了子网的概念，但仍然使用“掩码”这个词。

2011-8-31 《数据通信与计算机网络》——网络互连 28

**6.3 划分子网和构造超网**

CIDR记法的等效形式，省略点分十进制记法中低位的连续的0，如10.0.0.0/10简称为10/10，相当于指出其掩码是255.192.0.0。

也可以直接使用二进制记为

00001010 00xxxxxx xxxxxxxx xxxxxxxx

另一种简化记法是在网络前缀后加一个星号\*，如：

00001010 00\*

由于一个CIDR地址块可以表示许多（原来的网络）地址，所以在路由表中使用CIDR地址块查找目的网络，这样可以使路由表的一个项目可以表示很多个传统的分类地址的路由项目。这称为“路由聚合” (route aggregation)，也称为“构成超网” (supernetting)。路由聚合有利于减少路由器之间的路由选择信息的交换，从而提高了整个因特网的性能。

书上表6-7给出了常用的地址块。

2011-8-31 《数据通信与计算机网络》——网络互连 29

**6.3 划分子网和构造超网**

**CIDR 地址块划分举例**

有时把这种用CIDR分配的地址块称为一个“编址域”或“域”。用CIDR分配的地址块一定是2的整数幂。

从例子中可以看出，把四个系的路由聚合为大学的一个路由（即构成超网），是将网络前缀缩短。网络前缀越短，其地址块包含的地址数就越多。而在三级结构IP地址中，划分子网是使网络前缀变长。

2. 最长前缀匹配

使用CIDR时，路由表中的每个项目由“网络前缀”和“下一跳地址”组成。在查找路由表时可能会得到不止一个匹配结果。应选择哪一条呢？

因为前缀越长，地址块越小，因而路由就越具体（more specific），所以应该从匹配结果中选择具有最长网络前缀的匹配，这称为最长前缀匹配（longest-prefix matching）。

2011-8-31 《数据通信与计算机网络》——网络互连 30

### 6.3 划分子网和构造超网

若IP地址分配一开始就使用CIDR，那么可以按网络所在的地理位置分配地址块，从而大大减小路由表中的路由项目。

3. 使用二叉线查找路由表

当路由表的项目数很大时，如何设法减小路由表的查找时间是一个重要的问题。如线路速率10Gb/s，分组平均长度2000bit，那么路由器平均每秒要处理500万个分组（5Mpps）。

为了更有效地查找路由表，通常将无分类编址的路由表存放在一种层次的数据的结构中，然后自上而下按层次查找。常用的就是二叉线索（binary trie）。

2011-8-31 《数据通信与计算机网络》——网络互连 31

### 6.4 因特网控制报文协议ICMP

IP不保证可靠交付，为了提高IP数据报交付成功的机会，在网际层使用了因特网控制报文协议ICMP（Internet Control Message Protocol）。它允许主机或路由器报告差错和提供有关异常情况的报告。

ICMP不是高层协议，它和IP同处于网际层。

前4个字节都是一样的

0 8 16 31

类型 代码 检验和

(这4个字节取决于ICMP报文的类型)

ICMP的数据部分(长度取决于类型)

ICMP报文

首部 数据部分

IP数据报

2011-8-31 《数据通信与计算机网络》——网络互连 32

### 6.4 因特网控制报文协议ICMP

ICMP报文分为两种：ICMP差错报告报文和ICMP询问报文。

ICMP的前4个字节分三个字段：类型、代码和检验和。类型指出差错报告报文（5种）或询问报文（4种）的类型；代码字段进一步区分某类型的不同情况；检验和字段用来检验整个ICMP报文。

ICMP差错报告报文共有5种：

- (1) 终点不可达（3），分为网络不可达、主机不可达、协议不可达、端口不可达、需要分片但DF为1，源路由失败。
- (2) 源站抑制（4），用于拥塞时控制源站的发送速率。
- (3) 时间超过（11），收到TTL为0的数据报，丢弃后报告。
- (4) 参数问题（12），所收到的数据报首部中有的字段的值不正确。
- (5) 改变路由（5），路由器将改变路由由报文发给主机。

2011-8-31 《数据通信与计算机网络》——网络互连 33

### 6.4 因特网控制报文协议ICMP

ICMP差错报告报文中数据字段的格式

IP数据报的数据字段

收到的IP数据报

IP数据报首部 8字节

ICMP的前8字节

IP数据报首部 8字节

ICMP差错报告报文

装入ICMP报文的IP数据报

首部 ICMP差错报告报文

IP数据报

2011-8-31 《数据通信与计算机网络》——网络互连 34

### 6.4 因特网控制报文协议ICMP

以下几种情况不应发送ICMP差错报告报文：

- (1) 对ICMP差错报告报文不再发送；
- (2) 对第一个分片的数据报的所有后续数据报片都不再发送；
- (3) 对具有多播地址的数据报不发送；
- (4) 对具有特殊地址（如127.0.0.0）的数据报不发送。

ICMP询问报文有四种，即回送请求和回答、时间戳请求和回答、掩码地址请求和回答以及路由器询问和通告。

应用层的服务PING用来测试两个主机之间的连通性，它没有通过运输层而直接使用了ICMP的回送请求与回送回答报文。

时间戳请求与回答用来进行时钟同步和测量时间；主机使用掩码请求报文从子网掩码服务器得到某个接口的掩码地址；主机使用路由器询问和通告报文了解连接在本网络上的路由器是否正常工作。

2011-8-31 《数据通信与计算机网络》——网络互连 35

### 6.5 因特网的路由选择协议

#### 6.5.1 有关路由选择协议的几个基本概念

1. 理想的路由算法
 

在上一章已讨论过了。
2. 分层次的路由选择协议
 

因特网采用的路由选择协议主要是自适应的、分布式的、分层的路由选择协议。之所以要分层，主要是由于以下两个原因：

  - (1) 因特网规模非常大，现在已有几百万个路由器互连在一起。
  - (2) 许多接在因特网上的单位不希望外界了解本单位网络的布局及所使用的路由选择协议。

所以，因特网将整个互联网分为许多小的自治系统（autonomous system），简称AS。一个自治系统就是一个互联网，其最重要的特点就是自治系统有权决定在本系统内应采用何种路由选择协议。

2011-8-31 《数据通信与计算机网络》——网络互连 36

**6.5 因特网的路由选择协议**

一个自治系统内的所有路由器在本自治系统内都必须是连通的。

在此基础上，因特网把路由选择协议分为两大类：

(1) **内部网关协议IGP (Interior Gateway Protocol)**，即在一个自治系统内使用的路由选择协议。如RIP，OSPF。

(2) **外部网关协议EGP (External Gateway Protocol)**，位于不同自治系统的两个主机通信，当数据传到一个自治系统的边界时，就需要使用一种协议将路由选择信息传递到另一个自治系统中。这样的协议就是EGP。目前常用的是BGP-4。

名词的使用 (IRP, ERP)。

自治系统之间的路由选择也叫做**域间路由选择**，而自治系统内的路由选择又叫做**域内路由选择**。

2011-8-31 《数据通信与计算机网络》——网络互连 37

**6.5 因特网的路由选择协议**

**6.5.2 内部网关协议RIP**

1. 工作原理

路由信息协议RIP (Routing Information Protocol) 是一种分布式的基于距离向量的路由选择协议。最大优点是简单。

RIP协议要求网络中的每个路由器都要维护从它自己到其它每个目的网络的距离。RIP的距离即“跳数”，并规定一条路径最多只能经过15个路由器。可见RIP只适用于小型互联网。

RIP不能在两个网络之间同时使用多条路由。RIP选择一个具有最少路由器的路由，哪怕还存在另一条高速但路由器更多的路由。

分布式路由选择协议的特点是每个路由器都要不断地和其它路由器交换路由信息。所以必须弄清以下三个要点：在什么时候和哪些路由器交换什么信息？RIP的三个要点是：

2011-8-31 《数据通信与计算机网络》——网络互连 38

**6.5 因特网的路由选择协议**

(1) 按固定的时间间隔 (如30s) 交换路由信息。

(2) 仅和相邻路由器交换信息。相邻即在同一个网络上都有接口。

(3) 交换当前本路由器知道的全部信息，即自己的路由表。

RIP协议的收敛过程较快，收敛是指在自治系统中所有的结点都得到正确的路由选择信息 (最短距离、最小跳数) 的过程。

2. 距离向量算法

距离向量算法的要点：设X是结点A到B的最短路径上的一个结点。若将路径A→B拆成两段路径A→X和X→B，则每段路径 A→X和X→B也分别是结点A到X和结点X到B的最短路径。

RIP协议让互联网中的所有路由器都和自己的相邻路由器不断交换路由信息，并不断更新其路由表，使得从每个路由器到每个目的网络的路由都是最短的。具体的算法如下：

2011-8-31 《数据通信与计算机网络》——网络互连 39

**6.5 因特网的路由选择协议**

收到相邻路由器 (其地址为 X) 的一个 RIP 报文：

(1) 先修改此 RIP 报文中的所有项目：将“下一跳”字段中的地址都改为 X，并将所有的“距离”字段的值加 1。

(2) 对修改后的 RIP 报文中的每一个项目，重复以下步骤：  
若项目中的目的网络不在路由表中，则将该项目加到路由表中。  
否则 若下一跳字段给出的路由器地址是同样的，则将收到的项目替换原路由表中的项目。  
否则 若收到项目中的距离小于路由表中的距离，则进行更新，否则，什么也不做。

(3) 若 3 分钟还没有收到相邻路由器的更新路由表，则将此相邻路由器记为不可达的路由器，即将距离置为 16。

(4) 返回。

2011-8-31 《数据通信与计算机网络》——网络互连 40

**6.5 因特网的路由选择协议**

**RIP工作过程演示**

RIP协议用运输层的用户数据报UDP传送 (端口520)，因此RIP协议的位置应该在应用层。但转发IP数据报的过程是在网络层完成的。

3. RIP协议的报文格式

RIP2支持变长子网掩码和CIDR，提供简单的鉴别功能并支持多播。

RIP报文由首部 and 路由部分组成。首部的命令字段指出报文的意义。路由部分最多可包括25个路由，每个路由信息占20字节。

路由信息中的地址族标识符字段用来标志所使用的地址协议 (IP协议此字段填2)。路由标记填入自治系统的号码 (有可能收到自治系统外的路由选择信息)。

优点：简单，开销小。缺点：限制了规模；开销随规模增大而增加；坏消息传播得慢。

2011-8-31 《数据通信与计算机网络》——网络互连 41

**6.5 因特网的路由选择协议**

**6.5.3 内部网关协议OSPF**

1. OSPF协议的基本特点

开放最短路径优先OSPF (Open Shortest Path First) 是一种自适应的、分布式的链路状态协议 (link state protocol)。

OSPF的三个要点是：

(1) 当链路状态发生变化时交换信息。

(2) 通过洪泛向本自治系统中所有路由器发送信息。

(3) 发送的信息是与本路由器相邻的所有路由器的链路状态。

通过各路由器之间交换链路状态信息，所有路由器最终都能建立并维持一个链路状态数据库，实际上也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 (链路状态数据库同步)。每个路由器基于该数据库用Dijkstra算法构造自己的路由表。

2011-8-31 《数据通信与计算机网络》——网络互连 42

**6.5 因特网的路由选择协议**

OSPF的链路状态数据库能较快地更新，其更新过程收敛快。

为了使OSPF能用于规模很大的网络，OSPF支持把一个自治系统再划分为若干个更小的范围，叫作区域（Area）。OSPF使用层次结构的区域划分。

OSPF不用UDP而直接使用IP数据报传送（协议字段值89）。OSPF构成的数据报很短，好处是可减少路由信息通信量且不必分片。

OSPF的其它特点：

- (1) OSPF对不同类型的业务可计算出不同的路由。
- (2) 负载均衡。到同一目的网络有多条相同代价路径，分散通信量。
- (3) 所有在OSPF路由器间交换的分组都具有鉴别功能。
- (4) OSPF支持可变长度子网划分和CIDR。
- (5) 每个链路状态带有32bit序号，越大表示状态越新。

2011-8-31 《数据通信与计算机网络》——网络互连 43

**6.5 因特网的路由选择协议**

OSPF分组使用24字节固定长度首部。首部各字段的意义是：

版本：当前版本是2；类型，5种类型分组；

分组长度：包括首部，单位是字节；

路由器标识符：发送该分组的路由器的接口的IP地址；

区域标识符：分组属于的区域标识符；

检验和：检测分组中的差错；鉴别类型：0（不用）和1（口令）；

鉴别：类型为0填入0，类型为1填入8个字符的口令。

2. OSPF的五种分组类型

- (1) 问候（Hello）分组，发现和维持邻站可达性。
- (2) 数据库描述（Database Description）分组，向邻站发数据库摘要。
- (3) 链路状态请求（Link State Request）分组，请求邻站发详细信息。

2011-8-31 《数据通信与计算机网络》——网络互连 44

**6.5 因特网的路由选择协议**

- (4) 链路状态更新（Link State Update）分组，对全网更新链路状态。
- (5) 链路状态确认（Link State Acknowledgment）分组，对链路更新分组的确认。

OSPF规定，两个相邻路由器每隔10s交换一次问候分组。40s没收到问候分组，认为不可达。其它四种分组用于链路状态数据库的同步。

基本工作工程，问候发现邻站，相邻站交换摘要信息，请求自己缺少的信息，链路状态变化时用可靠的洪泛向全网广播更新链路状态信息。

为确保链路状态数据库与全网保持一致，OSPF规定每隔30分钟要刷新一次数据库中的链路状态。

当互联网规模很大时，OSPF协议比RIP好得多，且OSPF无“坏消息传得慢”的问题。

若N个路由器连接在一个以太网上，使用“指定的路由器”。

2011-8-31 《数据通信与计算机网络》——网络互连 45

**6.5 因特网的路由选择协议**

OSPF支持三种网络连接：两个路由器之间的点对点连接；具有广播功能的局域网；无广播功能的广域网。

**6.5.4 外部网关协议BGP**

边界网关协议BGP是不同自治系统的路由器之间交换路由信息的协议。

不同自治系统间为什么不能用RIP或OSPF？因特网规模太大、自治系统间路由选择非常困难；对自治系统间的路由选择，寻找最佳路由是很不现实的（自治系统之间只能交换可达性）；必须考虑政治、安全或经济方面的策略。

BGP力求寻找一条能到达目的网络且比较好的路由，并非要寻找一条最佳路由。BGP采用了路径向量（path vector）路由选择协议。

配置BGP时，每个自治系统都要选择至少一个路由器作为该自治系统的“BGP发言人”。

2011-8-31 《数据通信与计算机网络》——网络互连 46

**6.5 因特网的路由选择协议**

一个BGP发言人与其它自治系统的BGP发言人交换信息时先建立TCP连接（端口号179），然后交换BGP报文以建立BGP会话（session），利用BGP会话交换路由信息（如增加了新路由、撤销过时的路由及报告出差错的情况等）。通过TCP连接交换信息的两个发言人彼此称为对方的邻站或对等站。

BGP发言人互相交换了网络可达性信息（要到达某个网络所要经过的一系列自治系统）后，各发言人就根据所采用的策略从收到的信息中找出到达各自治系统的比较好的路由。

**BGP发言人交换路径向量**

BGP协议交换路由信息的结点数量级是自治系统数量的量级。

BGP支持CIDR。由于使用了路径向量的信息，可以很容易地避免产生兜圈子的路由。

2011-8-31 《数据通信与计算机网络》——网络互连 47

**6.5 因特网的路由选择协议**

BGP刚刚运行时，邻站交换整个BGP路由表。但以后只需要在发生变化时更新有变化的部分。

BGP-4使用四种报文：

- (1) 打开（Open）报文，与相邻的发言人建立关系。
- (2) 更新（Update）报文，发送某一路由信息，并列出来要撤销的路由。
- (3) 保活（Keepalive）报文，确认打开报文，周期性证实邻站关系。
- (4) 通知（Notification）报文，用来发送检测到的差错。

邻站每30s交换一次保活报文。更新报文一次可以撤销多条路由，但一个报文只能增加一条。

BGP报文的格式：

首部长19字节，分三个字段：标记（16字节）；长度（包括首部，单位是字节）；类型（1到4，对应四种报文）。

2011-8-31 《数据通信与计算机网络》——网络互连 48



XI'AN JIAOTONG UNIVERSITY

## 6.6 IP多播和因特网组管理协议

### 6.6.1 IP多播的基本概念

IP多播主要用于提供一对多的通信服务。 多播demo

PC机的网卡硬件可以判断收到帧的目的地址是否属于：本网卡的硬件地址（单播）；广播地址；组地址且本站已加入该组（多播）。

因特网上向多个目的站发送同样的数据报的方法：单播多次；多播。

采用多播协议可明显地减轻网络中各种资源的消耗。显然，在因特网范围的多播需要路由器的支持。

在因特网上进行IP多播的一些特点：

- (1) 多播使用组地址（D类），多播地址只能用于目的地址。
- (2) 永久组地址，由IANA指派。
- (3) 动态组成员；
- (4) 使用硬件进行多播。 硬件多播

2011-8-31      《数据通信与计算机网络》——网络互连      49

XI'AN JIAOTONG UNIVERSITY

## 6.6 IP多播和因特网组管理协议

### 6.6.2 因特网组管理协议IGMP

因特网组管理协议IGMP（Internet Group Management Protocol）位于网际层，用来帮助多播路由器识别加入到一个多播组的成员主机。与ICMP类似，IGMP使用IP数据报传递其报文。

IGMP可分为两个阶段：

- (1) 某个主机向多播组的多播地址发送IGMP报文，声明加入组。
- (2) 本地多播路由器通过周期性探询本地局域网维护动态的组成员。

主机在多播中的状态

每个主机都保留有记录其所加入的多播组的信息表。

IGMP的报文仅8个字节，共四个字段：**类型**（询问、成员关系报告、离开组），**响应时间**，**检验和**（与IP数据报计算方法相同）以及**组地址**。

2011-8-31      《数据通信与计算机网络》——网络互连      50

XI'AN JIAOTONG UNIVERSITY

## 6.6 IP多播和因特网组管理协议

### 6.6.3 多播路由选择

IGMP协议规定了主机如何与本地的多播路由器进行交互，但并没有指明在多播路由器之间应如何交换组成员关系的信息。

多播路由选择比单播路由选择要复杂得多。这主要是因为：

- (1) 动态的组成员导致即使网络拓扑不变，多播路由也会不断变化；
- (2) 多播路由选择要确定什么时候需要将多播数据报复制并转发副本；
- (3) 多播数据报可以由非多播组的成员产生。

因特网上第一个多播路由选择协议是**距离向量多播路由选择协议**。

当多播数据报在传输过程中遇到不运行多播软件的路由器或网络时，就要采用**隧道技术（tunneling）**。 ▶

还提出多种多播路由选择协议：开放最短通路优先的多播扩展，核心基干树，协议无关多播-稀疏方式，协议无关多播-密集方式。

2011-8-31      《数据通信与计算机网络》——网络互连      51

XI'AN JIAOTONG UNIVERSITY

## 6.7 虚拟专用网和网络地址转换

### 6.7.1 虚拟专用网VPN

一个机构能申请到的IP地址有限，而且并非机构内的所有计算机都需要接入因特网，所以在机构内部可以自行分配IP地址。为解决地址的二义性问题，就需要使用**专用地址（private address）**。

专用地址只能用于一个机构的内部通信，不能用于和因特网上的主机通信。RFC 1918指明的专用地址是：

- (1) 10.0.0.0到10.255.255.255（10/8），即一个A类地址。
- (2) 172.16.0.0到172.31.255.255（172.16/12），即16个B类地址。
- (3) 192.168.0.0到192.168.255.255（192.168/16），即256个连续的C类地址。

使用专用IP地址的互连网络称为**专用互连网或本地互连网**，简称**专用网**。专用IP地址也叫做**可重用地址（reusable address）**。

2011-8-31      《数据通信与计算机网络》——网络互连      52

XI'AN JIAOTONG UNIVERSITY

## 6.7 虚拟专用网和网络地址转换

一个很大的机构的许多部门分布在相距很远的一些地点，在每个地点都有自己的专用网。如果这些专用网需要经常通信，有两种方法：一是租用电信专线；二是利用因特网实现本机构的专用网，这样的专用网称为**虚拟专用网VPN（Virtual Private Network）**。专用网的数据报通过因特网时使用隧道技术。

一个机构内部建立的虚拟专用网又称为**内联网（Intranet）**；不同机构之间共同建立的虚拟专用网又称为**外联网（Extranet）**。

### 6.7.2 网络地址转换NAT

专用网内部的一些主机需要和因特网上的主机进行通信，怎么办呢？最简单的办法是再申请一些全球IP地址。另一种方法是使用**网络地址转换NAT（Network Address Translation）**。

仅用IP进行转换；同时使用IP和运输层的端口号进行转换。

2011-8-31      《数据通信与计算机网络》——网络互连      53

XI'AN JIAOTONG UNIVERSITY

## 6.8 下一代的网际协议IPv6

### 6.8.1 解决IP地址耗尽的措施

IPv4地址将要耗尽，解决办法：采用无分类编址CIDR；采用网络地址转换NAT方法；采用IPv6。

### 6.8.2 IPv6的基本首部

IPv6的主要变化有：

- (1) 更大的地址空间。从32bit到128bit。
- (2) 扩展的地址层次结构。
- (3) 灵活的首部格式。提高路由器的处理效率。
- (4) 改进的选项。IPv4的选项是固定不变的。
- (5) 允许协议继续扩充。IPv4的功能是固定不变的。
- (6) 支持即插即用（即自动配置）。
- (7) 支持资源的预分配。

2011-8-31      《数据通信与计算机网络》——网络互连      54

### 6.8 下一代的网际协议IPv6

IPv6首部长度固定为40字节，称**基本首部**。取消了IPv4中不必要的功能，取消了首部检验和，加快了路由器处理分组的速度。

IPv6在基本首部后允许有0到多个扩展首部，但扩展首部不属于数据报首部，它和数据合起来叫做数据报的有效载荷（payload）。

**IPv6基本首部**

#### 6.8.3 IPv6的扩展首部

##### 1. 扩展首部及下一个首部字段

IPv6把IPv4首部中选项的功能都放在扩展首部中，数据报途经的路由器只处理扩展首部中的逐跳选项扩展首部，从而大大提高了处理速率。

IPv6定义了六种扩展首部：**逐跳选项；路由选择；分片；鉴别；封装安全有效载荷；目的站选项**。使用多个首部时应按以上先后顺序出现。

2011-8-31 《数据通信与计算机网络》——网络互连 55

### 6.8 下一代的网际协议IPv6

#### 2. 扩展首部举例

以分片扩展首部为例进行说明。

IPv6将分片限制为由源站来完成。源站可以采用保证的最小MTU（1280字节），或在发送数据前使用路径最大传送单元发现（Path MTU Discovery）确定最小的MTU。

分片扩展首部的格式如下：

片偏移（13bit），指明本片在原数据报中的偏移量，单位8个字节。  
M=1表示后面还有分片；M=0表示是最后一片。标识符。

比特	0	8	16	29	31
	下一个首部	保留	片 偏 移	保留	M
	标 识 符				

2011-8-31 《数据通信与计算机网络》——网络互连 56

### 6.8 下一代的网际协议IPv6

采用端到端的分片可以减小路由器开销，但改变了因特网的基本假设，不再能容易地改变路由。IPv6允许中间的路由器采用隧道技术来传送太长的数据报。

#### 6.8.4 IPv6的地址空间

##### 1. 128bit的地址空间

IPv6的目的地址可以是单播（unicast）、多播（multicast）或任播（anycast）。

IPv6将实现IPv6协议的主机和路由器均称为结点，并将IPv6地址分配给结点上面的接口。一个接口可以有多个单播地址。一个结点接口的单播地址可以用来惟一地标志该结点。

为了地址分配和修改的方便，IPv6允许给一个给定的网络指派多个前缀，也允许对一个主机的给定接口同时指派多个地址。

2011-8-31 《数据通信与计算机网络》——网络互连 57

### 6.8 下一代的网际协议IPv6

IPv6地址的记法：点分十进制、冒号十六进制记法（colon hexadecimal notation）。每16bit用十六进制表示，中间用冒号隔开。

冒号十六进制支持零压缩（zero compression），即一连串的零可以用一对冒号代替，规定任一地址只能用一次零压缩。如  
FF05:0:0:0:0:0:B3 可以写为：FF05::B3。

冒号十六进制可结合有点分十进制记法的后缀，这在IPv4向IPv6过渡的转换阶段很有用。如：  
0:0:0:0:0:0:128.10.2.1，  
再使用零压缩可以写为：  
::128.10.2.1

CIDR的斜线记法仍然可用。

2011-8-31 《数据通信与计算机网络》——网络互连 58

### 6.8 下一代的网际协议IPv6

#### 2. 地址空间的分配

IPv6把地址空间分为长度可变的两部分，分别是类型前缀（3到10比特）和其余部分。类型前缀定义了地址的目的。

地址的分配主要考虑：如何与IPv4兼容；如何使路由器更快地查找路由；如何对各种接口的硬件地址进行直接编码。

前缀00000000B是保留的一小部分用于和IPv4兼容的。

为了路由器更快地查找路由，IPv6扩展了地址的分级概念，使用以下三个等级：

- (1) 第一级（顶级），指明全球都知道的公共拓扑。
- (2) 第二级（地点级），指明单个地点。
- (3) 第三级，指明单个的网络接口。

全球唯一地址格式EUI-64。

2011-8-31 《数据通信与计算机网络》——网络互连 59

### 6.8 下一代的网际协议IPv6


#### 6.8.5 从IPv4向IPv6过渡

过渡方法只能采用逐步演进的办法，并向后兼容。

双协议栈（dual stack）

2011-8-31 《数据通信与计算机网络》——网络互连 60

## 6.8 下一代的网际协议IPv6

  
XI'AN JIAOTONG  
UNIVERSITY

另一种方法是使用隧道技术。 用隧道技术过渡

### 6.8.6 ICMPv6

ICMPv6的前4个字节与ICMPv4相同，从第5字节开始是报文主体。  
取消了使用很少的报文。

ICMPv6的报文分为两类：差错报文（error message）和提供信息的报文（informational message）。

ICMPv6规定差错报文的类型字段的最高位是0，而提供信息的报文类型字段的最高位是1。

ICMPv6用IPv6封装，其前面的一个首部中的“下一个首部字段”的值是58。IPv4协议字段为1时表示数据部分是ICMP报文。

2011-8-31《数据通信与计算机网络》——网络互连61