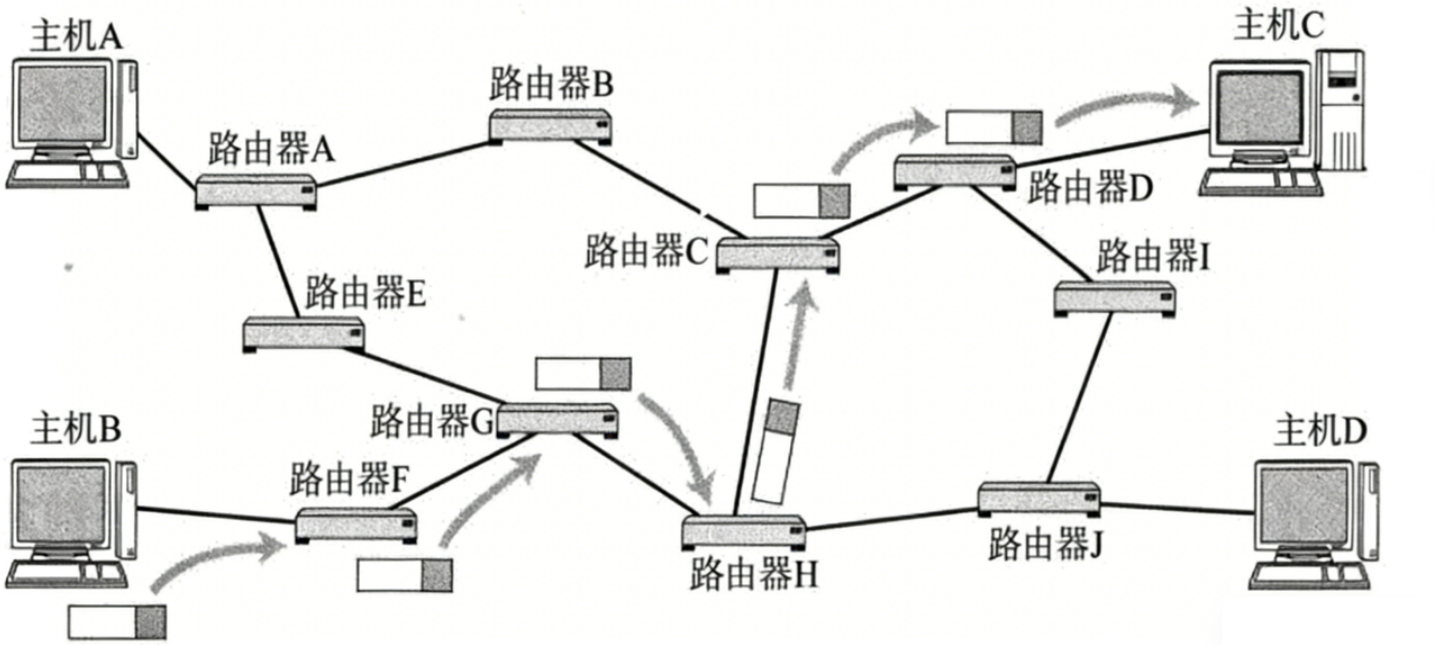
网络层

网络层

在复杂的网络环境中确定一个合适的路径.

IP协议



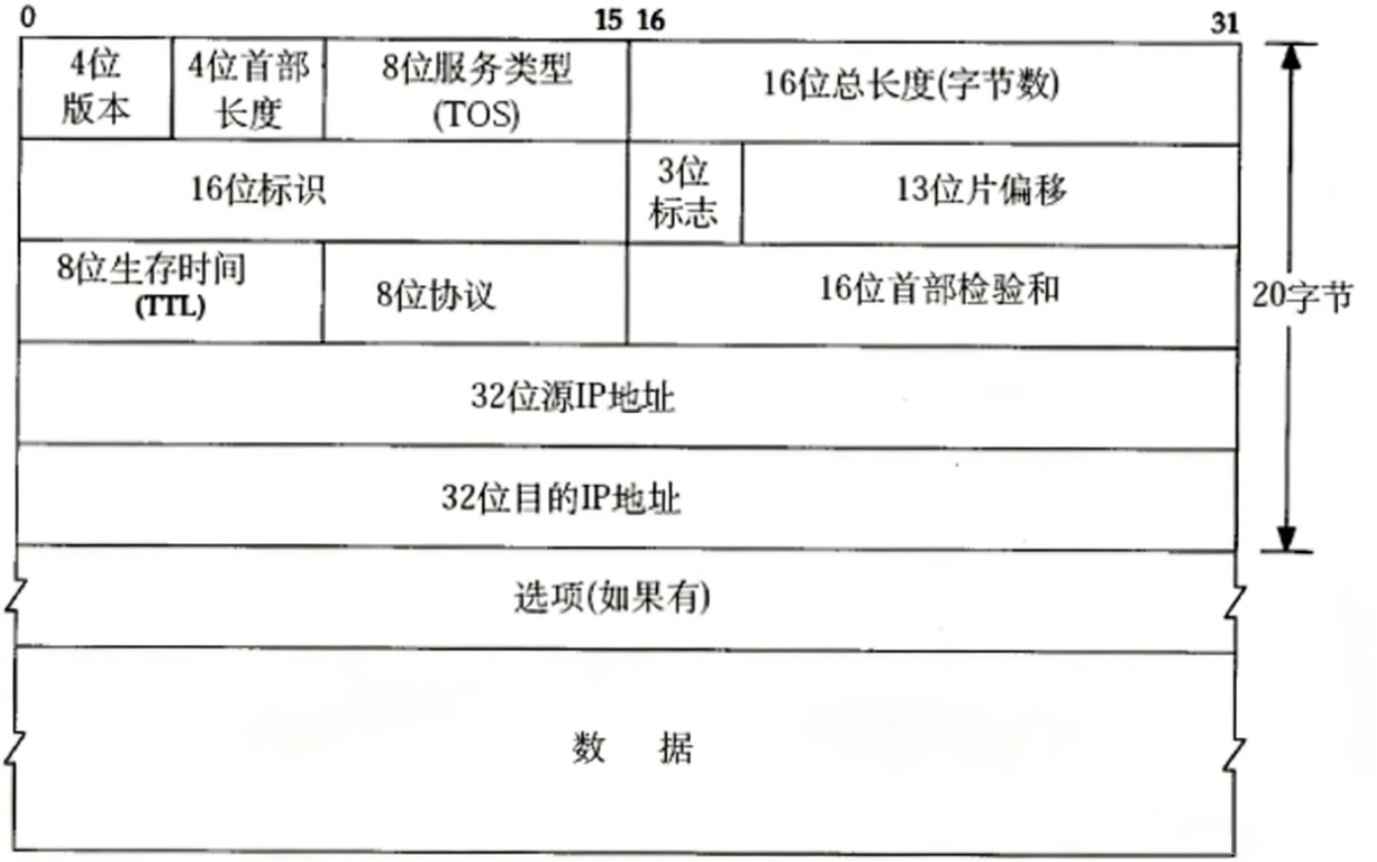
基本概念

主机: 配有IP地址, 也要进行路由控制的设备;

路由器: 即配有IP地址, 又能进行路由控制;

节点: 主机和路由器的统称;

协议头格式



· 4位版本号(version): 指定IP协议的版本, 对于IPv4来说, 就是4.

· *4位头部长度(header length): IP头部的长度是多少个32bit, 也就是 length* 4 的字节数. 4bit表示最大的数字是15, 因此IP头部最大长度是60字节.

· 8位服务类型(Type Of Service): 3位优先权字段(已经弃用), 4位TOS字段, 和1位保留字段(必须置为0). 4位TOS分别表示: 最小延时, 最大吞吐量, 最高可靠性, 最小成本. 这四者相互冲突, 只能选择一个. 对于ssh/telnet这样的应用程序, 最小延时比较重要; 对于ftp这样的程序, 最大吞吐量比较重要.

· 16位总长度(total length): IP数据报整体占多少个字节.

· 16位标识(id): 唯一的标识主机发送的报文. 如果IP报文在数据链路层被分片了, 那么每一个片里面的这个id都是相同的.

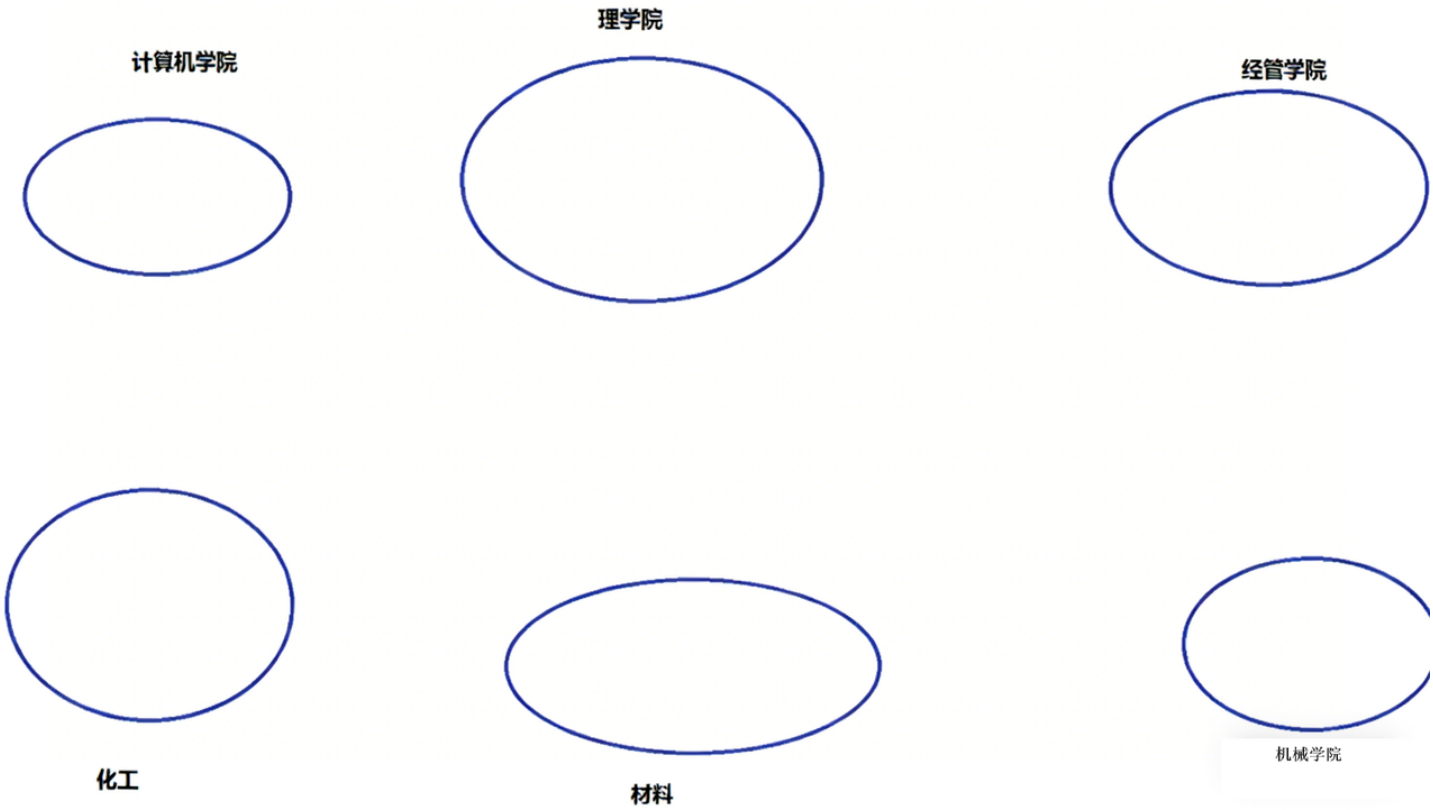
· 3位标志字段: 第一位保留(保留的意思是现在不用, 但是还没想好说不定以后要用到). 第二位置为1表示禁止分片, 这时候如果报文长度超过MTU, IP模块就会丢弃报文. 第三位表示"更多分片", 如果分片了的话, 最后一个分片置为0, 其他是1. 类似于一个结束标记.

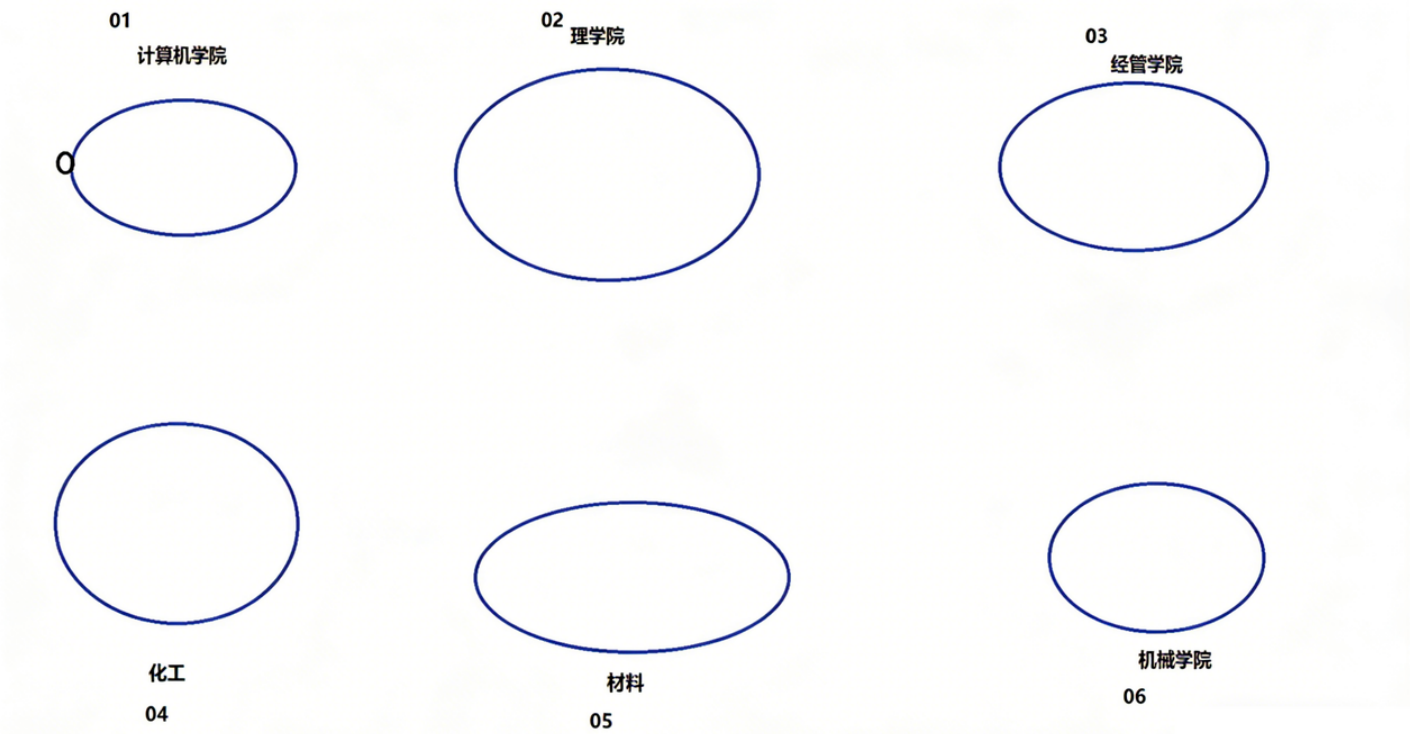
· 13位分片偏移(framegament offset): 是分片相对于原始IP报文开始处的偏移. 其实就是在表示当前分片在原报文中处在哪个位置. 实际偏移的字节数是这个值 8 得到的. 因此, 除了最后一个报文之外, 其他报文的长度必须是8的整数倍(否则报文就不连续了).

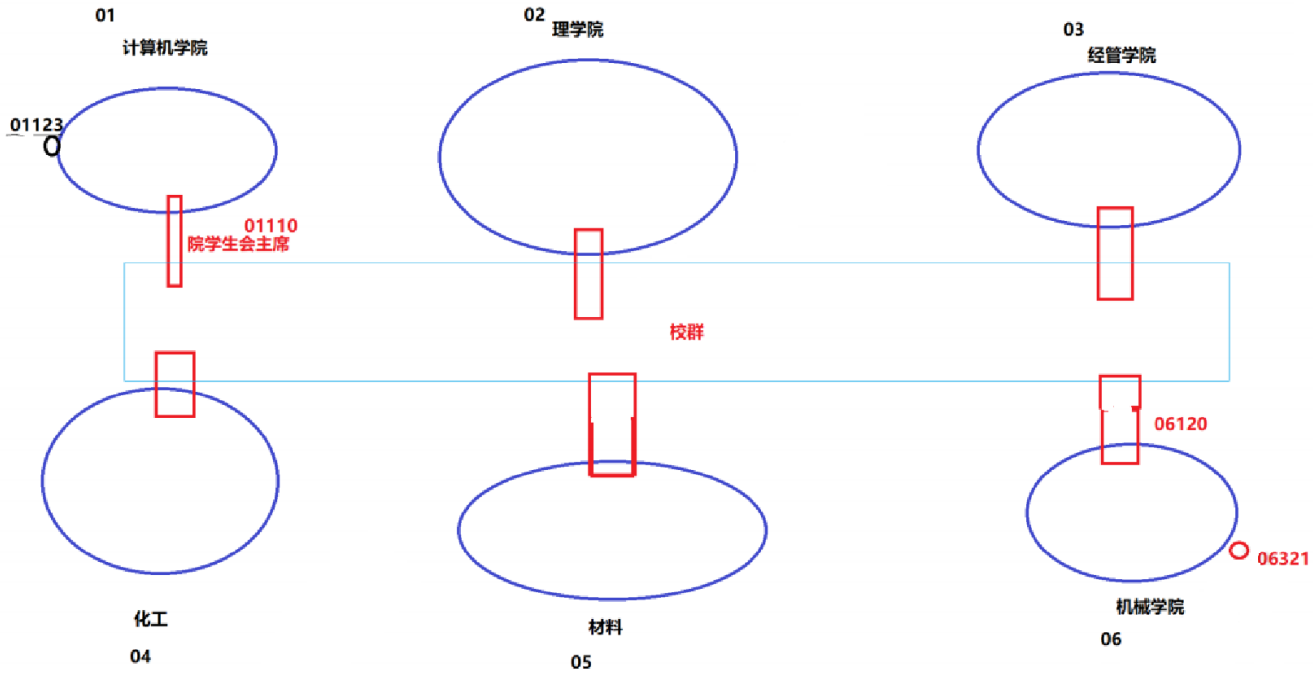
· 8位生存时间(Time To Live, TTL): 数据报到达目的地的最大报文跳数. 一般是64. 每次经过一个路由, TTL -= 1, 一直减到0还没到达, 那么就丢弃了. 这个字段主要是用来防止出现路由循环

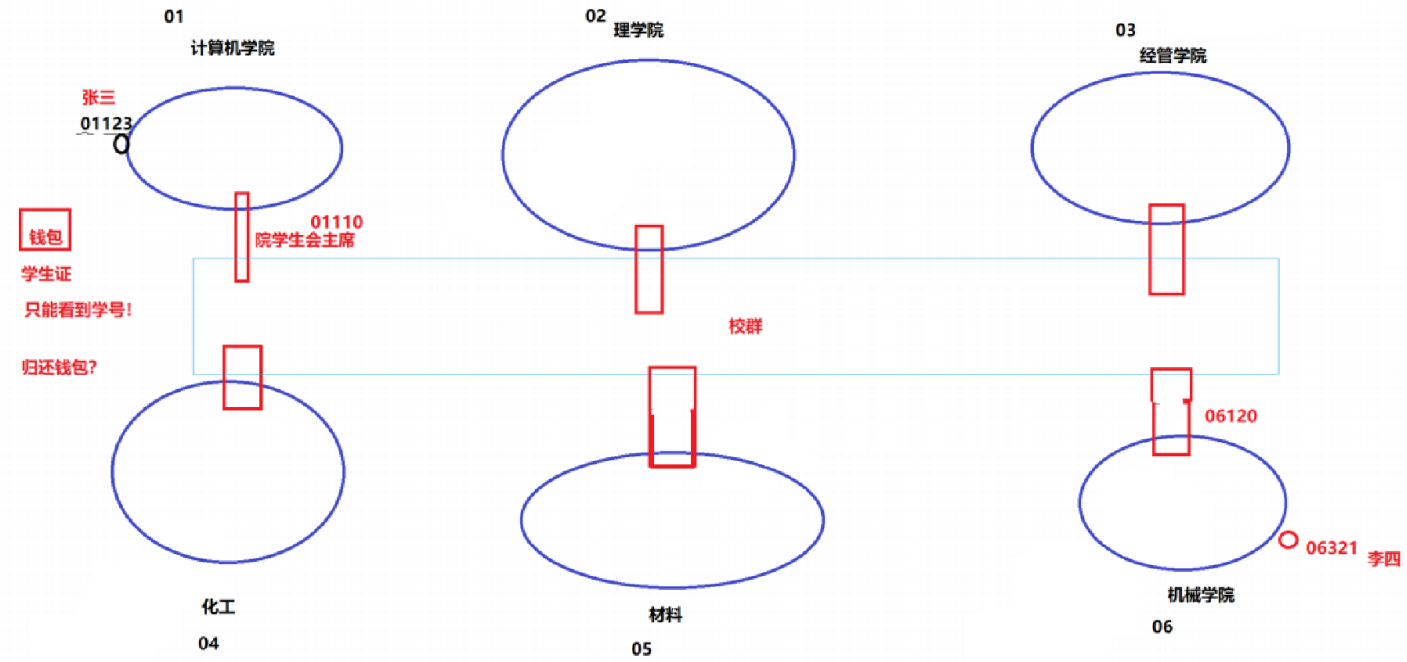
· 8位协议: 表示上层协议的类型

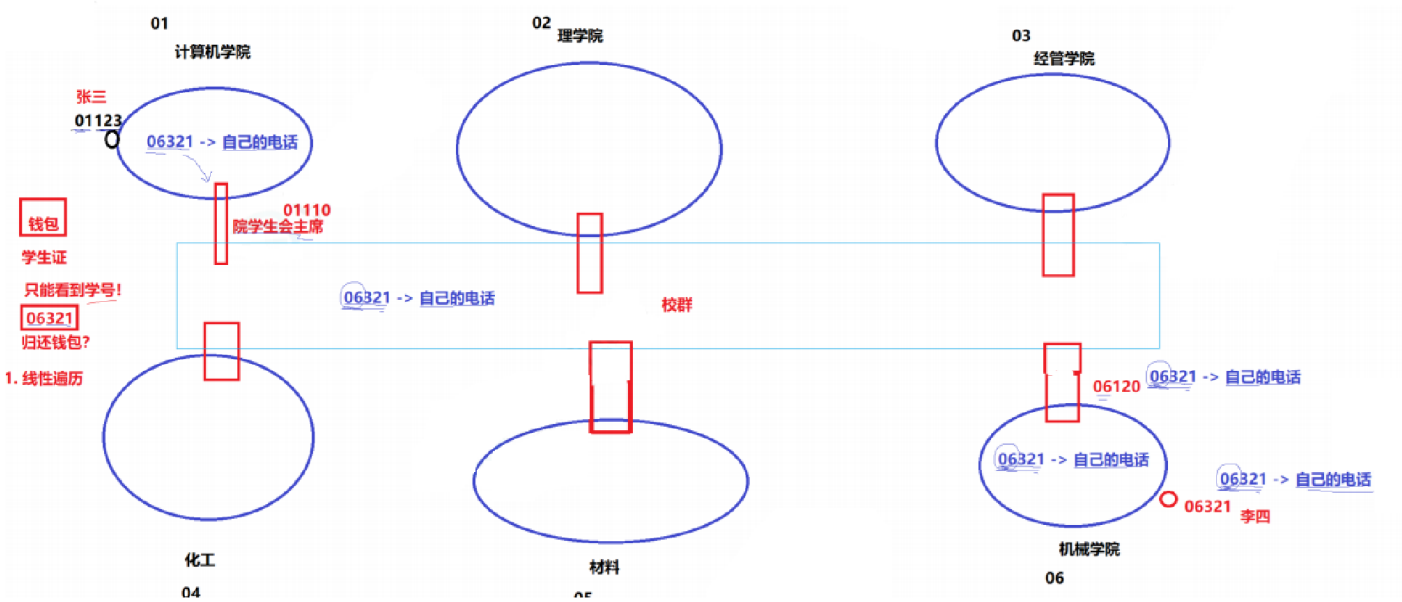
- 16位头部校验和: 使用CRC进行校验, 来鉴别头部是否损坏.
- 32位源地址和32位目标地址: 表示发送端和接收端.
- 选项字段(不定长, 最多40字节): 略






01
计算机学院
01123
01110
院学生会主席
02
理学院
03
经管学院
校群
06120
06321
化工
04
材料
05
机械学院
06

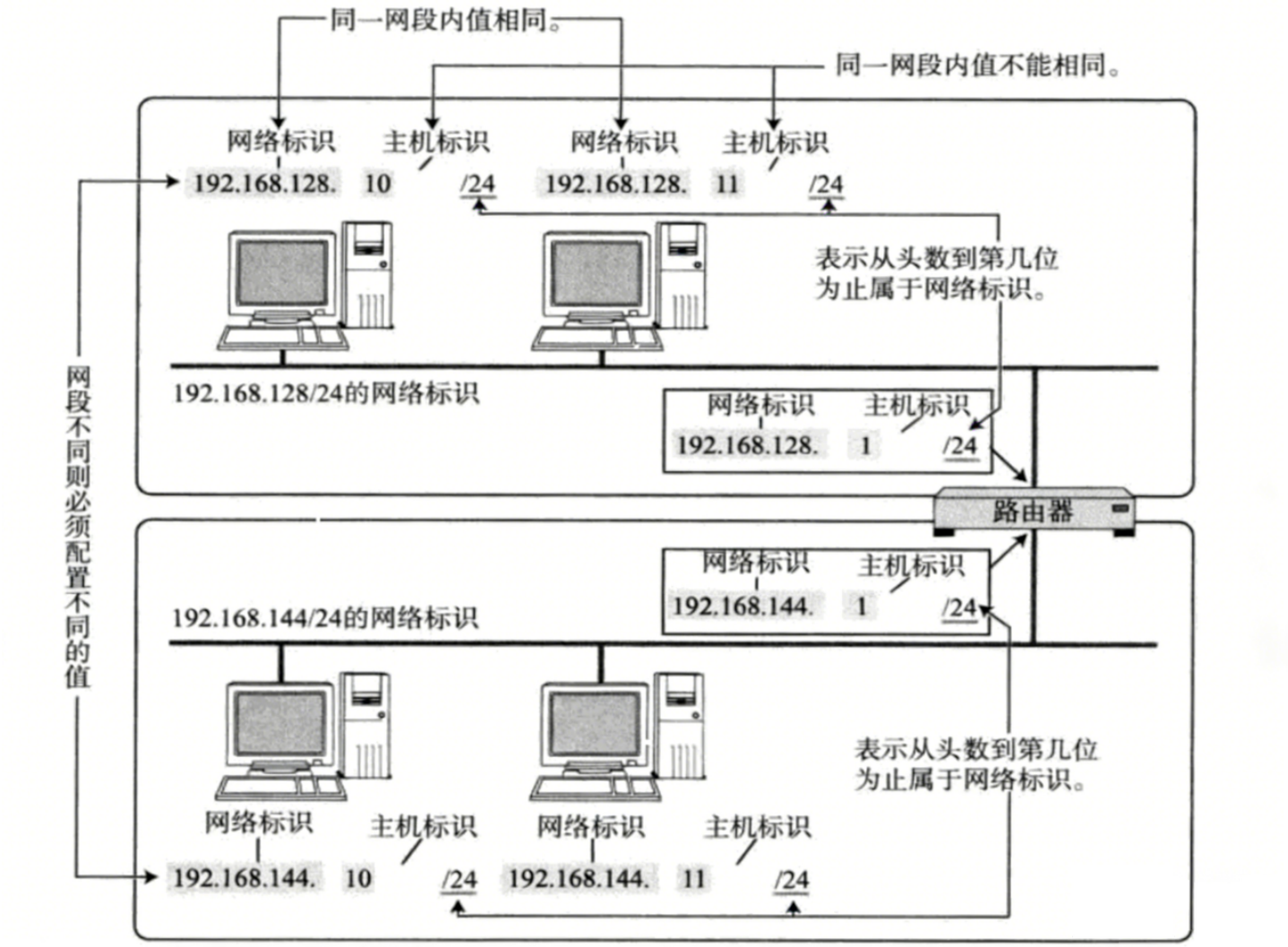

01
计算机学院
张三
01123
钱包
学生证
只能看到学号!
归还钱包?
01110
院学生会主席
02
理学院
03
经管学院
校群
06120
06321
李四
化工
04
材料
05
机械学院
06



网段划分(重要)

IP地址分为两个部分, 网络号和主机号

> · 网络号: 保证相互连接的两个网段具有不同的标识;
>
> · 主机号: 同一网段内, 主机之间具有相同的网络号, 但是必须有不同的主机号;



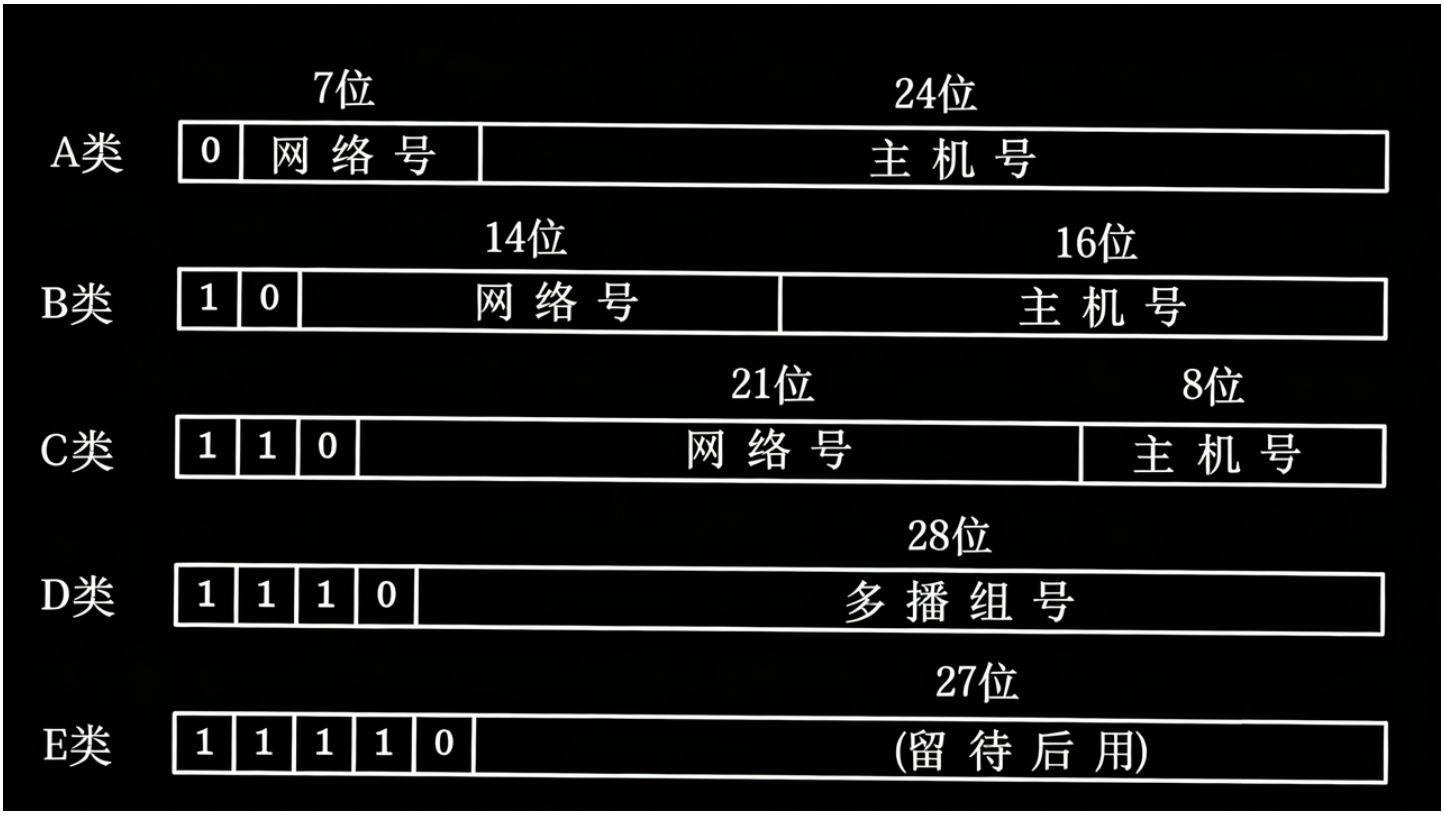
> · 不同的子网其实就是把网络号相同的主机放到一起.
>
> · 如果在子网中新增一台主机, 则这台主机的网络号和这个子网的网络号一致, 但是主机号必须不能和子网中的其他主机重复.

通过合理设置主机号和网络号, 就可以保证在相互连接的网络中, 每台主机的IP地址都不相同.

那么问题来了, 手·动管理子网内的IP, 是一个相当麻烦的事情.

> · 有一种技术叫做DHCP, 能够自动的给子网内新增主机节点分配IP地址, 避免了手·动管理IP的不便.
>
> · 一般的路由器都带有DHCP功能. 因此路由器也可以看做一个DHCP服务器

过去曾经提出一种划分网络号和主机号的方案, 把所有IP 地址分为五类, 如下图所示(该图出 自[TCPIP])。



> · A类 0.0.0.0到127.255.255.255
>
> · B类 128.0.0.0到191.255.255.255
>
> · C类 192.0.0.0到223.255.255.255
>
> · D类 224.0.0.0到239.255.255.255
>
> · E类 240.0.0.0到247.255.255.255

随着Internet的飞速发展,这种划分方案的局限性很快显现出来,大多数组织都申请B类网络地址, 导致B类地址很快就分配完了, 而A类却浪费了大量地址;

> · 例如, 申请了一个B类地址, 理论上一个子网内能允许6万5千多个主机. A类地址的子网内的主机数更多.
>
> · 然而实际网络架设中, 不会存在一个子网内有这么多的情况. 因此大量的IP地址都被浪费掉了.

针对这种情况提出了新的划分方案, 称为CIDR(Classless Interdomain Routing)(无类别域间路由):

> · 引入一个额外的子网掩码(subnet mask)来区分网络号和主机号;
> · 子网掩码也是一个32位的正整数. 通常用一串 "0" 来结尾;
> · 将IP地址和子网掩码进行 "按位与" 操作, 得到的结果就是网络号;
> · 网络号和主机号的划分与这个IP地址是A类、B类还是C类无关;

下面举两个例子:

划分子网的例子1

IP地址	140.252.20.68	8C FC 14 44
子网掩码	255.255.255.0	FF FF FF 00
网络号	140.252.20.0	8C FC 14 00
子网地址范围	140.252.20.0~140.252.20.255	

划分子网的例子2

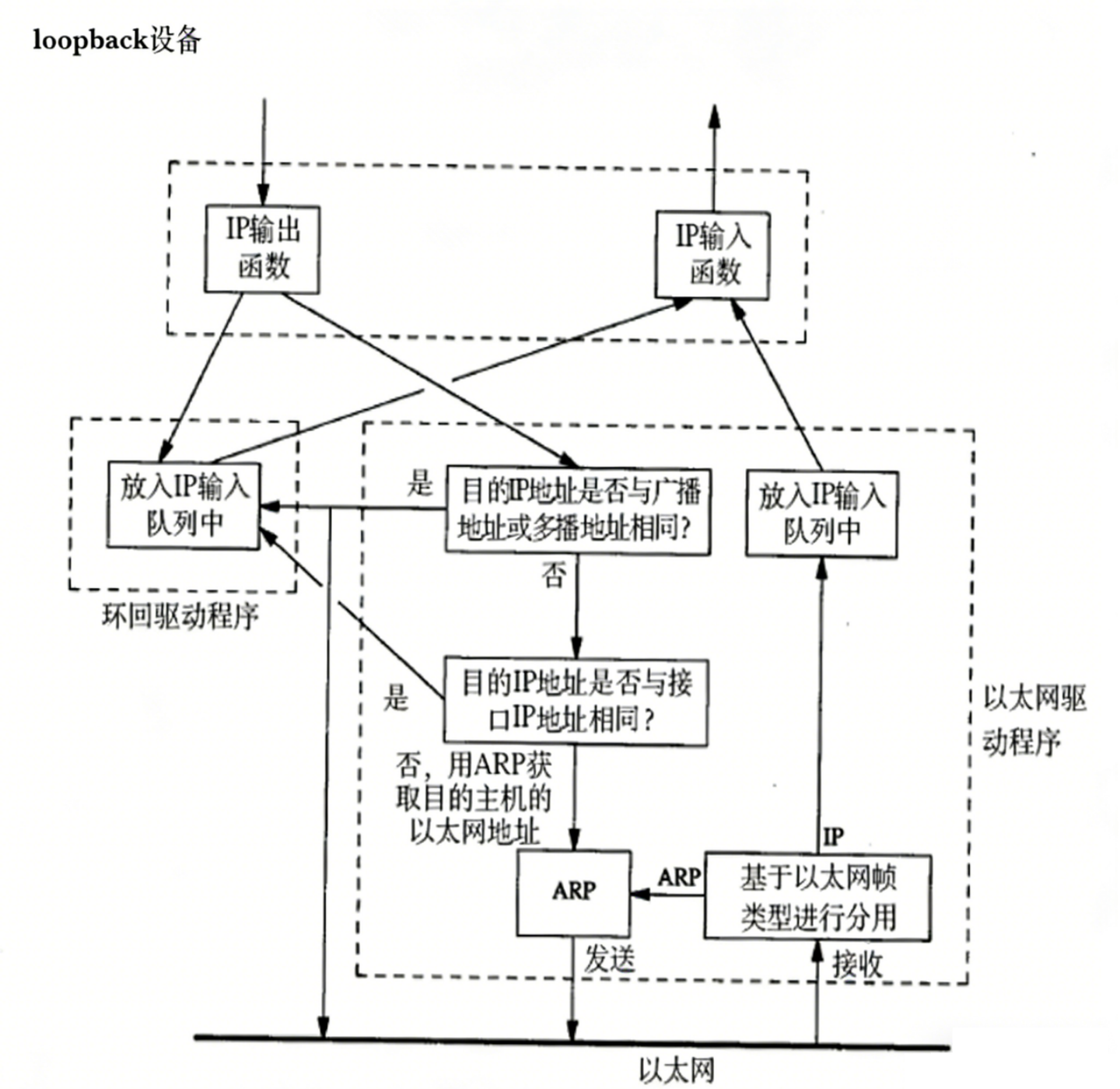
IP地址	140.252.20.68	8C FC 14 44
子网掩码	255.255.255.240	FF FF FF F0
网络号	140.252.20.64	8C FC 14 40
子网地址范围	140.252.20.64~140.252.20.79	

可见,IP地址与子网掩码做与运算可以得到网络号, 主机号从全0到全1就是子网的地址范围;

IP地址和子网掩码还有一种更简洁的表示方法,例如140.252.20.68/24,表示IP地址为140.252.20.68, 子网掩码的高24位是1,也就是255.255.255.0

特殊的IP地址

> · 将IP地址中的主机地址全部设为0, 就成为了网络号, 代表这个局域网;
> · 将IP地址中的主机地址全部设为1, 就成为了广播地址, 用于给同一个链路中相互连接的所有主机发送数据包;
> · 127.*的IP地址用于本机环回(loop back)测试,通常是127.0.0.1



IP地址的数量限制

我们知道, IP地址(IPv4)是一个4字节32位的正整数. 那么一共只有 2的32次方 个IP地址, 大概是43亿左右. TCP/IP协议规定, 每个主机都需要有一个IP地址.

这意味着, 一共只有43亿台主机能接入网络么?

实际上, 由于一些特殊的IP地址的存在, 数量远不足43亿; 另外IP地址并非是按照主机台数来配置的, 而是每一个网卡都需要配置一个或多个IP地址.

CIDR在一定程度上缓解了IP地址不够用的问题(提高了利用率, 减少了浪费, 但是IP地址的绝对上限并没有增加), 仍然不是很够用. 这时候有三种方式来解决:

· 动态分配IP地址: 只给接入网络的设备分配IP地址. 因此同一个MAC地址的设备, 每次接入互联网中, 得到的IP地址不一定是相同的;

· NAT技术(后面会重点介绍);

· IPv6: IPv6并不是IPv4的简单升级版. 这是互不相干的两个协议, 彼此并不兼容; IPv6用16字节128位来表示一个IP地址; 但是目前IPv6还没有普及;

私有IP地址和公网IP地址

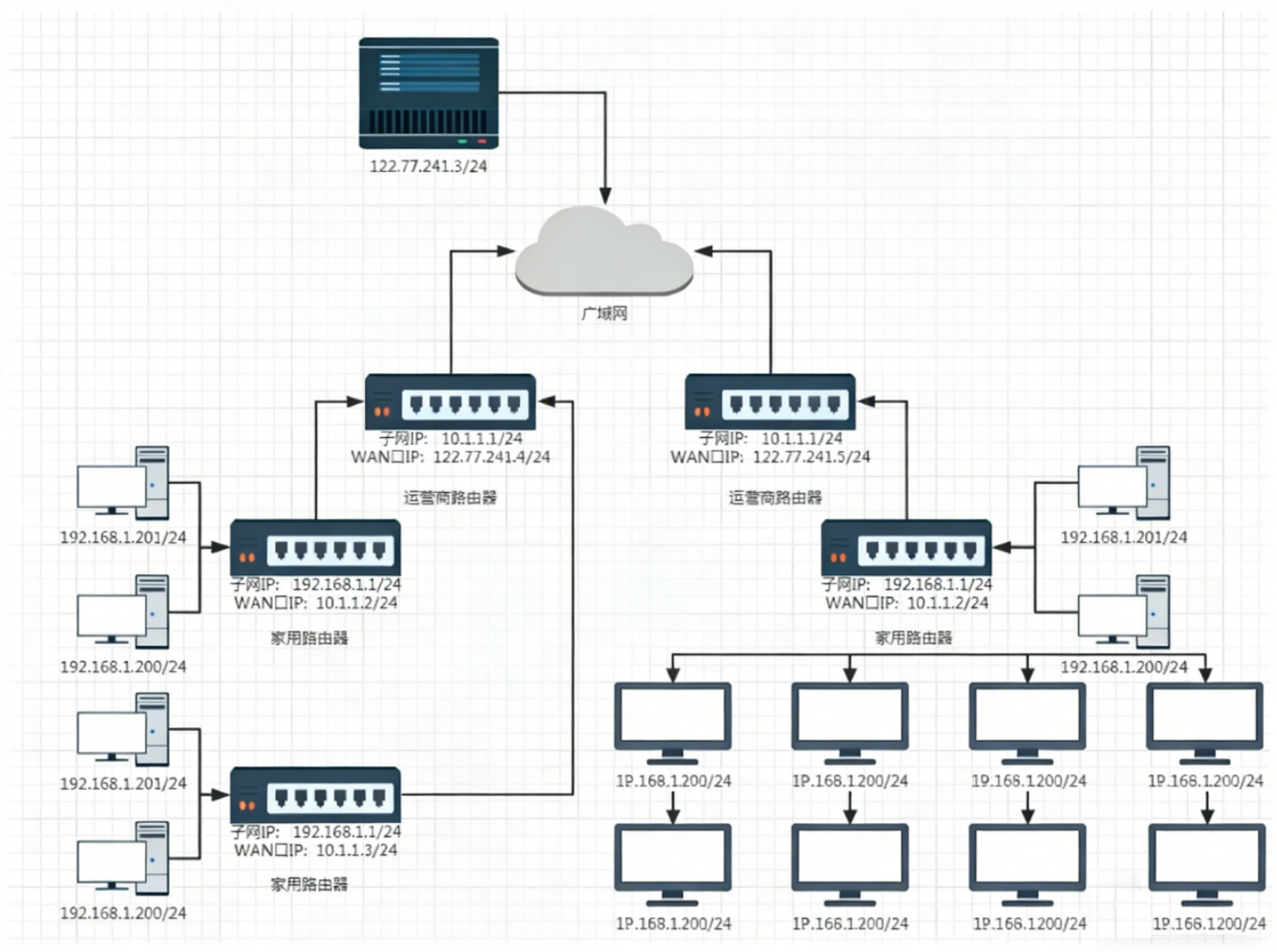
如果一个组织内部组建局域网,IP地址只用于局域网内的通信,而不直接连到Internet上,理论上 使用任意的IP地址都可以,但是RFC 1918规定了用于组建局域网的私有IP地址

· 10.*,前8位是网络号,共16,777,216个地址

· 172.16.*到172.31.*,前12位是网络号,共1,048,576个地址

· 192.168.*,前16位是网络号,共65,536个地址

包含在这个范围中的, 都成为私有IP, 其余的则称为全局IP(或公网IP);



· 一个路由器可以配置两个IP地址, 一个是WAN口IP, 一个是LAN口IP(子网IP).

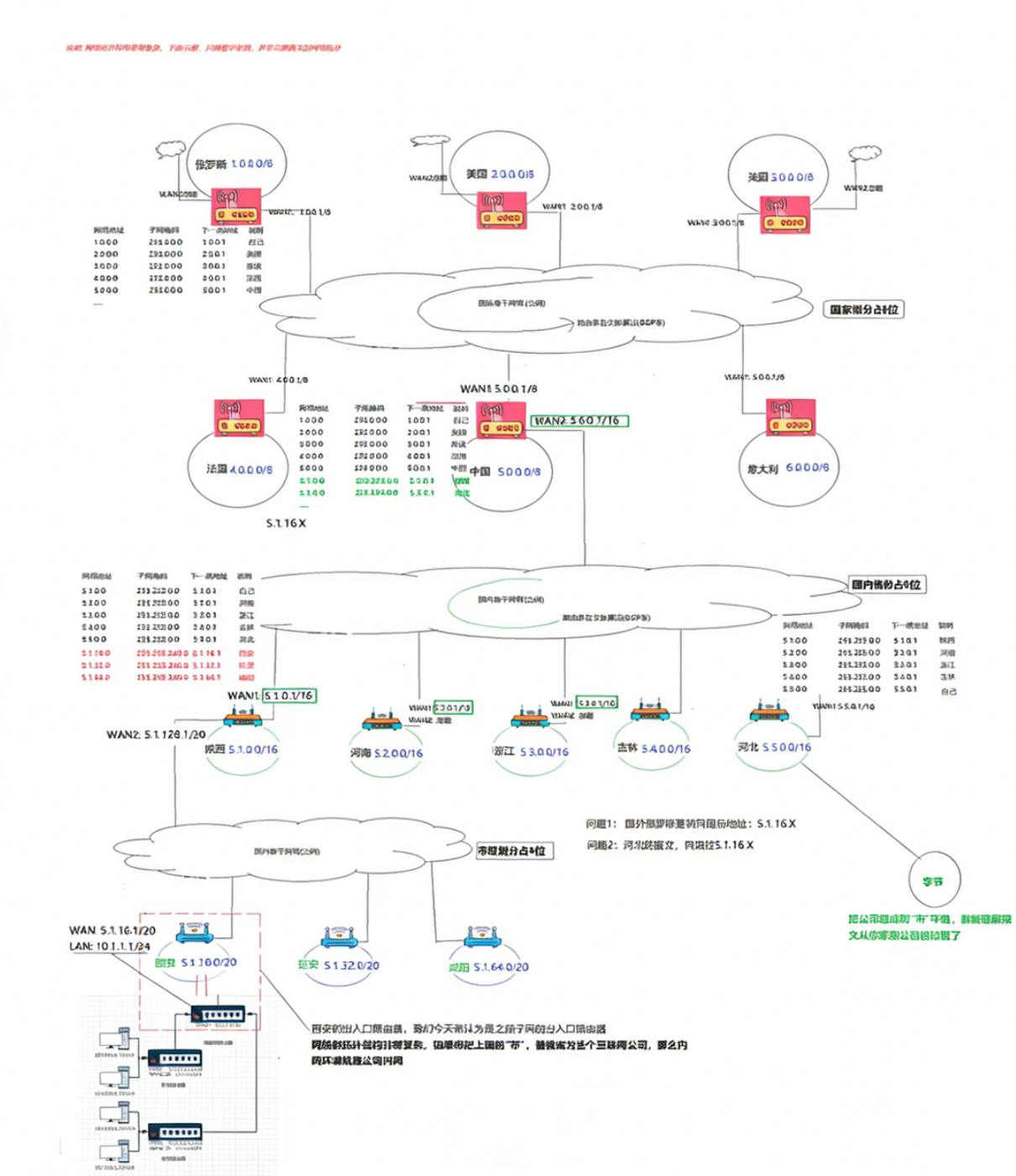
· 路由器LAN口连接的主机, 都从属于当前这个路由器的子网中.

> · 不同的路由器, 子网IP其实都是一样的(通常都是192.168.1.1). 子网内的主机IP地址不能重复. 但是子网之间的IP地址就可以重复了.
>
> · 每一个家用路由器, 其实又作为运营商路由器的子网中的一个节点. 这样的运营商路由器可能会有很多级, 最外层的运营商路由器, WAN口IP就是一个公网IP了.
>
> · 子网内的主机需要和外网进行通信时, 路由器将IP首部中的IP地址进行替换(替换成WAN口IP), 这样逐级替换, 最终数据包中的IP地址成为一个公网IP. 这种技术称为NAT(Network Address Translation，网络地址转换).
>
> · 如果希望我们自己实现的服务器程序, 能够在公网上被访问到, 就需要把程序部署在一台具有外网IP的服务器上. 这样的服务器可以在阿里云/腾讯云上进行购买.

地区或者国家IP的分布：https://zh-hans.ipshu.com/country-list

尝试理解公网

> 真实的网络结构非常复杂，即涉及到划分公网IP的组织，ICANN，还要在全球范围内进行区域划分，比如亚太，北美，欧洲等，又要考虑各个国家内部的ISP代理，整体拓扑非常复杂，我们简化所有过程，简单理解公网即可
>
> 附录有关于建设公网的参与者，和职责说明，AI提示词：公网的整个构建过程，需要涉及到谁，各自核心角色和作用是什么?



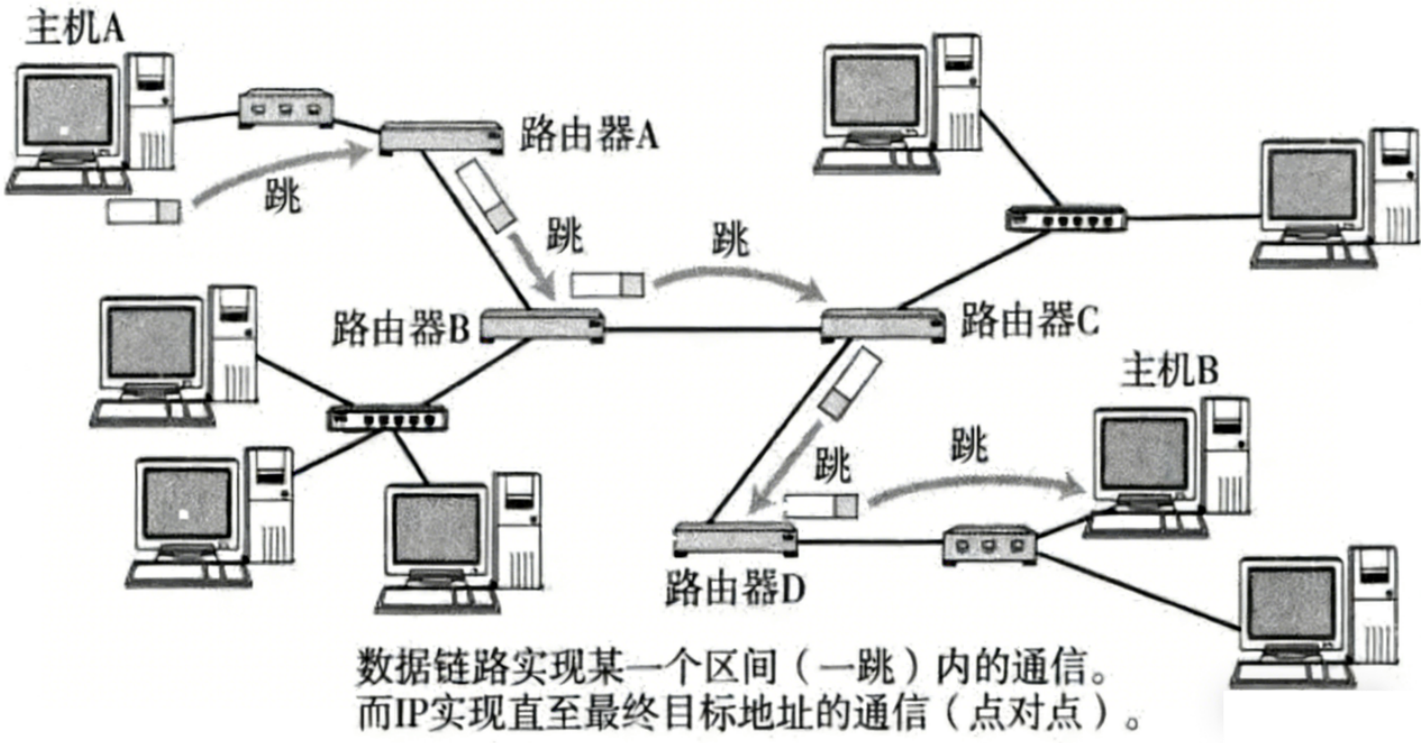
路由

在复杂的网络结构中, 找出一条通往终点的路线;

[唐僧问路例子1]

路由的过程, 就是这样一跳一跳(Hop by Hop) "问路" 的过程.

所谓 "一跳" 就是数据链路层中的一个区间. 具体在以太网中指从源MAC地址到目的MAC地址之间的帧传输区间.

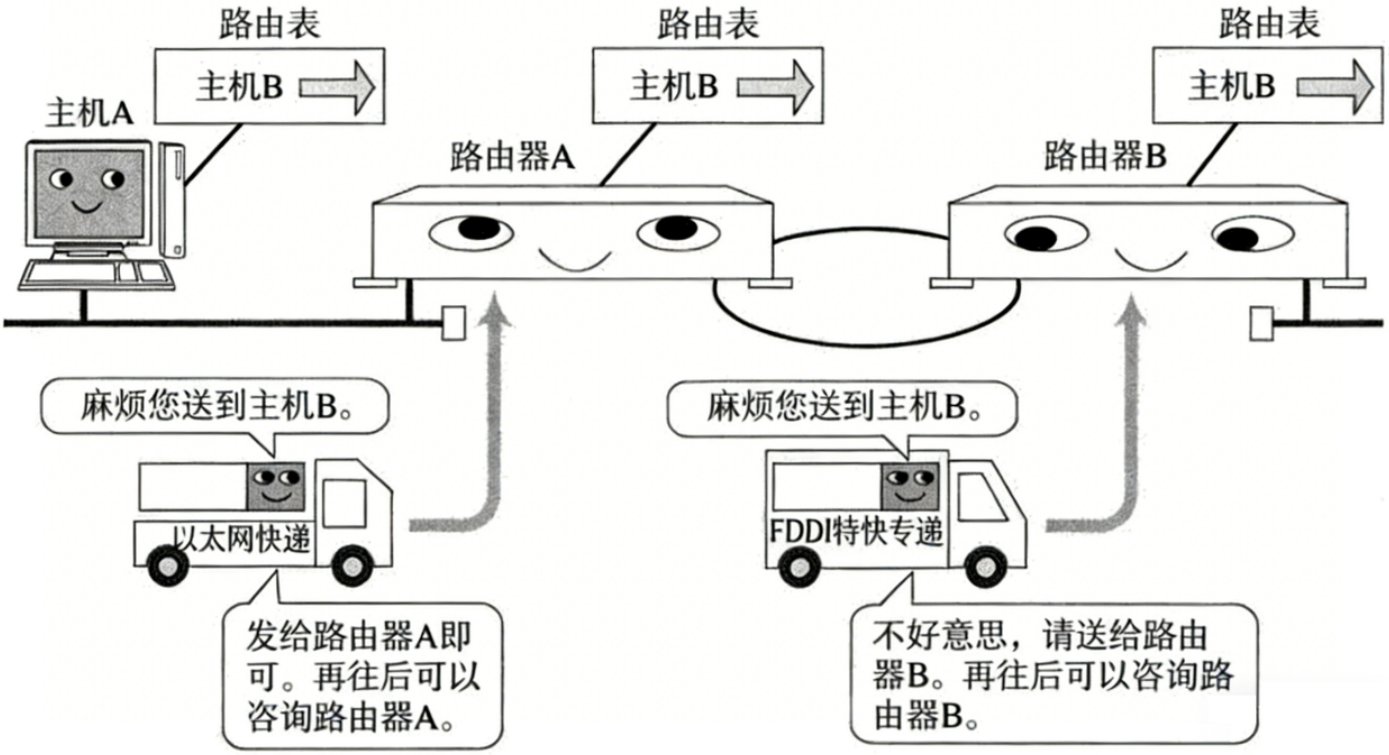


IP数据包的传输过程也和问路一样.

> · 当IP数据包, 到达路由器时, 路由器会先查看目的IP;
> · 路由器决定这个数据包是能直接发送给目标主机, 还是需要发送给下一个路由器;
> · 依次反复, 一直到达目标IP地址;

那么如何判定当前这个数据包该发送到哪里呢? 这个就依靠每个节点内部维护一个路由表;

[唐僧问路例子2]



> · 路由表可以使用route命令查看
>
> · 如果目的IP命中了路由表, 就直接转发即可;
>
> · 路由表中的最后一行,主要由下一跳地址和发送接口两部分组成,当目的地址与路由表中其它行都不匹配时,就按缺省路由条目规定的接口发送到下一跳地址。

假设某主机上的网络接口配置和路由表如下:

```
Destination     Gateway         Genmask         Flags Metric Ref
Use Iface
192.168.10.0    *               255.255.255.0   U     0      0
0 eth0
192.168.56.0    *               255.255.255.0   U     0      0
0 eth1
127.0.0.0       *               255.0.0.0       U     0      0
0 lo
default         192.168.10.1    0.0.0.0         UG    0      0
0 eth0
```

> · 这台主机有两个网络接口,一个网络接口连到192.168.10.0/24网络,另一个网络接口连到192.168.56.0/24网络;
>
> · 路由表的Destination是目的网络地址,Genmask是子网掩码,Gateway是下一跳地址,Iface是发送接口,Flags中的U标志表示此条目有效(可以禁用某些 条目),G标志表示此条目的下一跳地址是某个路由器的地址,没有G标志的条目表示目的网络地址是与本机接口直接相连的网络,不必经路由器转发;

转发过程例1: 如果要发送的数据包的目的地址是192.168.56.3

> · 跟第一行的子网掩码做与运算得 到192.168.56.0,与第一行的目的网络地址不符

> · 再跟第二行的子网掩码做与运算得 到192.168.56.0,正是第二行的目的网络地址,因此从eth1接口发送出去;
>
> · 由于192.168.56.0/24正 是与eth1 接口直接相连的网络,因此可以直接发到目的主机,不需要经路由器转发;

> 转发过程例2: 如果要发送的数据包的目的地址是202.10.1.2
>
> · 依次和路由表前几项进行对比, 发现都不匹配;
>
> · 按缺省路由条目, 从eth0接口发出去, 发往192.168.10.1路由器;
>
> · 由192.168.10.1路由器根据它的路由表决定下一跳地址;

路由表生成算法(选学)

路由表可以由网络管理员手·动维护(静态路由), 也可以通过一些算法自动生成(动态路由).

请同学们课后自己调研一些相关的生成算法, 例如距离向量算法, LS算法, Dijkstra算法等.

附录:

以下是公网构建全流程的参与主体及工作职责表格:

表格

	阶段	参与主体	核心职责	典型成果案例
1	物理基础设施	设备制造商（华为/思科）	生产光缆/路由器/服务器	单模光纤衰减 ≤0.18dB
2		能源供应商	保障数据中心7×24供电	Tier IV级数据中心 （99.
3	协议标准	IETF	制定TCP/IP/HTTP等协议	RFC 793（TCP协议规范
4		IEEE	制定802.3/802.11标准	100G以太网标准（802.
5	地址分配	ICANN	全球IP/域名根管理	IPv6 ::/12地址块分配
6		APNIC/RIPE NCC	区域IP分配与回收	中国203.0.113.0/24地址
7	骨干网络	Tier1运营商（中国电信）	跨洋光缆建设与维护	亚太直达海缆（108Tbp
8		IXP（DE-CIX）	提供网络交换节点	法兰克福交换中心峰值
9	接入服务	ISP（Comcast/中国联通）	部署FTTH/5G接入网	XGS-PON万兆光纤入户
10	监管合规	工信部/	频谱分配与网络政策制定	中国IPv6普及率 35%（
11		CERT	网络安全事件响应	Log4j漏洞全球应急处理
12	监管合规	AWS/阿里云	提供云计算与CDN服务	全球边缘节点超2300个
13		Netflix/抖音	内容分发与流量优化	自适应码率（ABR）技术
14	终端连接	手机厂商（苹果/小米）	实现协议栈与天线优化	WiFi6实测速率1.2Gbps

14 条记录