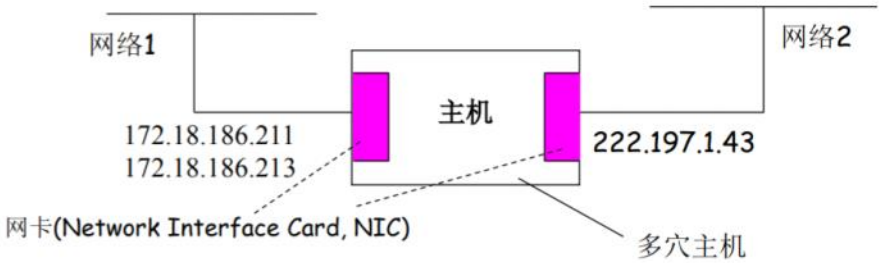
IP地址和ARP协议

2019年4月14日 9:51

IP地址空间

- 48位的MAC地址和32位的IP地址都是全局的(全球分配)，但是IP地址空间是分层的，是可路由的(routable)。
- IP地址由ICANN统一负责并逐级分配。亚洲由APNIC负责，中国由CNNIC负责。
- IP地址属于接口(网卡) 。主机或路由器的每个接口可以配置一个或多个IP地址。



ICANN --The Internet Corporation for Assigned Names and Numbers

IP地址结构

- 一个IP地址可以划分为两个部分: 网络号(network numbers)和主机号(host identifier)。
- 网络号也称为网络前缀(network prefix)、网络标识 (network ID) 。它是用来确定拥有该IP地址的主机位于哪个网络，而主机号用于确定属于该网络的哪台主机。

有类网

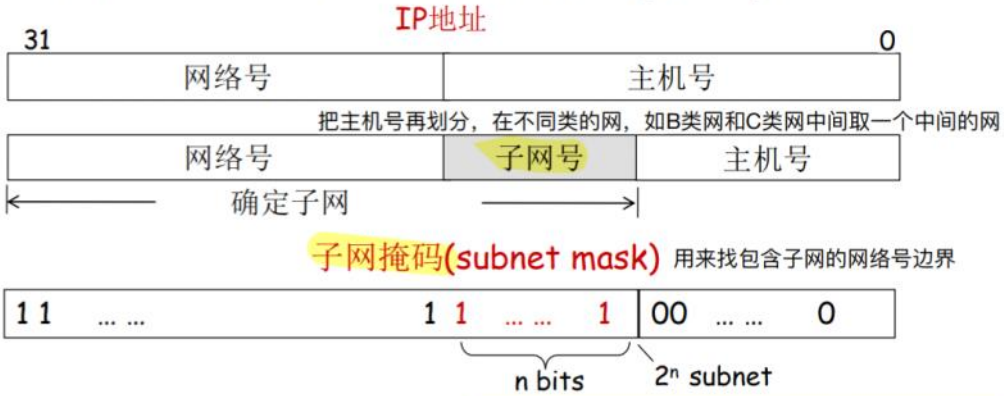
点分十进制(dotted decimal)

* 实际上，在有类网模型下，可用的A类网个数要减2，因为网络号全0和全1不可用。B类网和C类网个数也要减2。

在网络号全0或者全1的情况下一般不可用，所以个数要减2

子网划分

- 一个有类网可以划分为多个相同大小的子网(subnet)：



子网掩码也可以用点分十进制表示：C类网192.168.1.0划分为四个子网的子网掩码为255.255.255.192，子网号分别为00、01、10、11。

同一子网的子网掩码相同
用子网掩码和地址做与运算，如果网络号相同，则是在同一个子网

主机号为全1或者全0的地址被保留，不能使用。

* 子网号为全0或全1的子网现在都可以使用（以前规定不能使用）。

在一般的C类网中，可以连接256台主机。但用子网划分，每个子网都是相同的数量，如果想要划分成数量不同的子网，就需要使用变长子网划分。

变长子网掩码

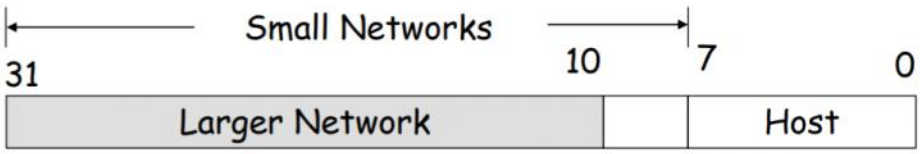
- 变长子网掩码(Variable-Length Subnet Mask，VLSM) 允许把一个有类网划分为多个不同大小的子网。
- 例如，给定一个有类网199.1.12.0，如何把它们划分为四个子网，使它们可以分别容纳 100，60，25和20台主机?

子网1：199.1.12.0XXXXXXX/25　(100台主机)
子网2：199.1.12.10XXXXXX/26　(60台主机)
子网3：199.1.12.110XXXXX/27　(25台主机)
子网4：199.1.12.111XXXXX/27　(10台主机)

用长度来表示子网掩码：/26表示255.255.255.192

除了有类网，还有无类的网

无类域间路由选择协议



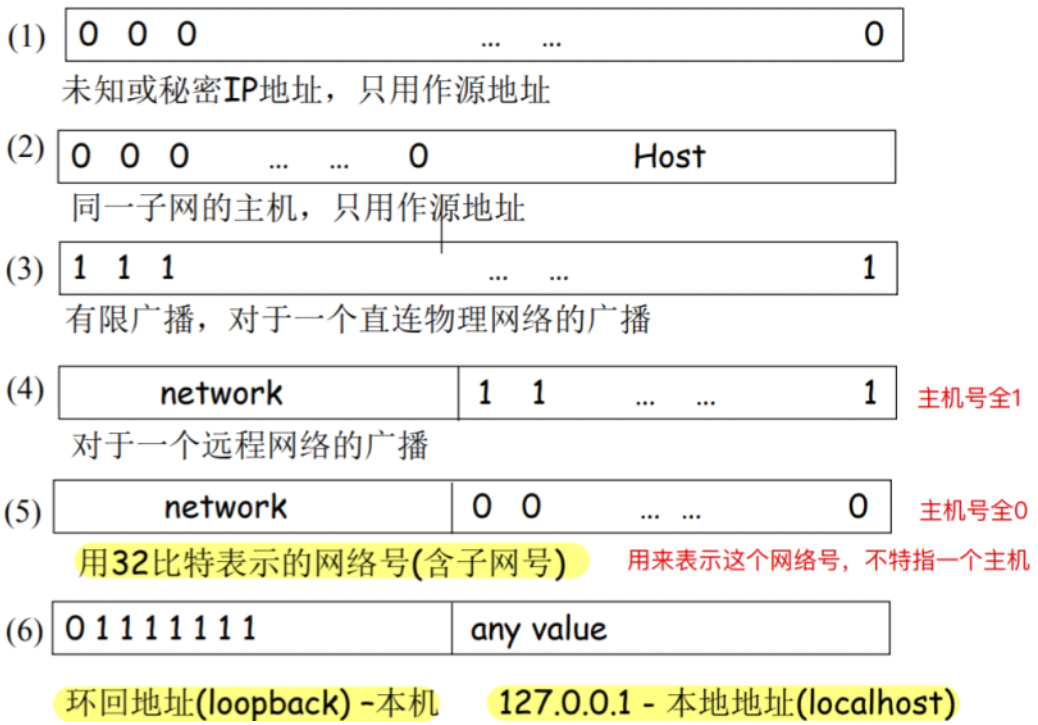
- 无类域间路由选择协议(Classless Inter-Domain Routing，CIDR)允许把多个有类网合并为一个更大的网络，称为超网(supernet)。
- 例如，把有类网192.24.8.0~192.24.15.0合并为网络号为192.24.8.0、子网掩码为255.255.248.0的超网。
- CIDR可以显著减少路由表中路由的数量，例如，上例就把八个路由减少路由减少为一个路由，称为路由聚合(route aggregation).
- 通过引入CIDR，加上子网掩码，现在的网络号(可能包含子网号)可以看成是没有边界，即是无类的。

合并8个C类网

CIDR可以减少路由表中的路由数量，缩小路由表的尺寸。

在计算机网络中，**路由表**（routing table）或称**路由**择域信息库（RIB, Routing Information Base），是一个存储在路由器或者联网计算机中的电子表格（文件）或类数据库。 **路由表**存储着指向特定网络地址的路径（在有些情况下，还记录有路径的**路由**度量值）。

特殊的IP地址



广播分组的目标IP地址的主机部分全为1，这意味着本地网络（广播域）中的所有主机都将接收并查看该分组。诸如ARP和DHCP等很多网络协议都使用广播。

例如：

C类网络192.168.1.0的默认子网掩码为255.255.255.0（掩码的255个数对应网络的网络地址个数），其广播地址为192.168.1.255，其主机部分为十进制数255或二进制数11111111（全为1）；

B类网络172.16.0.0的默认子网掩码为255.255.0.0，其广播地址为172.16.255.255；

A类网络10.0.0.0的默认子网掩码为255.0.0.0，其广播地址为10.255.255.255。

注意：主机号全0和全1不可用

注意：全1的有限广播不会被路由器转发，会被路由器过滤 ， 但远程网络的广播会被转发

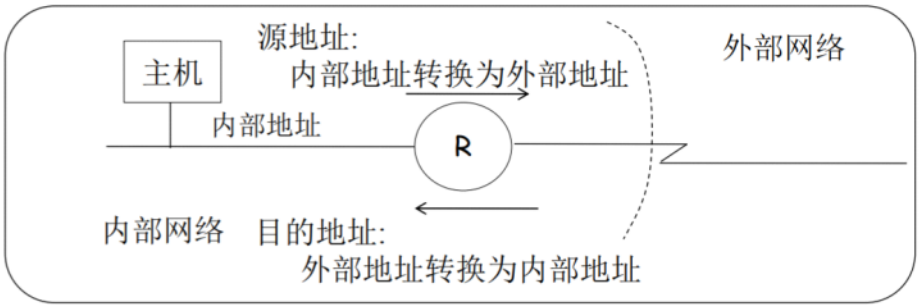
私有IP地址

- 私有IP地址就是无需IANA分配、任何人都可以使用的IP地址：
  （1）10.0.0.0 ~ 10.255.255.255
  （2）172.16.0.0 ~ 172.31.255.255
  （3）192.168.0.0 ~ 192.168.255.255
- 私有地址只能用于内部网络。主干网上的路由器会过滤掉目的地址为私有地址的IP数据报。因此，离开内部网络的IP数据报必须使用由IANA分配的全局地址作为目的地址。

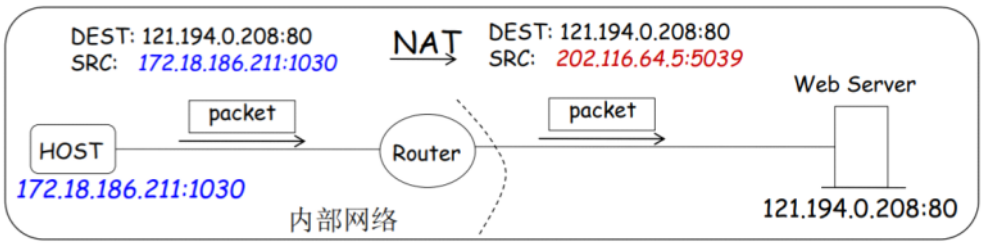
分别对应A、B、C类网

网络地址转换

- 网络地址转换(Network Address Translation，NAT)是一种把内部地址映射为外部地址的技术。例如，把私有地址映射为全局地址。
- 出口路由器在内网数据报发往外网时自动把内网地址映射为外网地址的方法称为动态NAT。每个动态映射都关联一个TTL。如果在TTL时间内没有使用一个映射，该映射将被出口路由器删除。直接由管理员加入映射的方法称为静态NAT。静态NAT加入的映射不会被自动删除。



- NAPT(Network Address Port Translation) 把端口号也加入到NAT的映射中，也称为PAT(Port Address Translation)或过载NAT(NAT with overload)。
- 下面的图显示了从私有网络的主机发包给Web服务器的地址转换方法：



source address	destination address	global address	src port	dest. port	global port	prot ocol	connection released	timer
172.18.186.211	121.194.0.208	202.116.64.5	1030	80	5309	TCP	no	2.0

* The timer expires in 2 minutes

- **网络地址转换**又称**网络掩蔽**、**IP掩蔽**（英语：**N**etwork **A**ddress **T**ranslation，缩写：**NAT**），在计算机网络中是一种在IP数据包通过路由器或防火墙时重写来源IP地址或目的IP地址的技术。这种技术被普遍使用在有多台主机但只通过一个公有IP地址访问因特网的私有网络中。它是一个方便且得到了广泛应用的技术。当然，NAT也让主机之间的通信变得复杂，导致了通信效率的降低。
- NAT是作为一种解决IPv4地址短缺以避免保留IP地址困难的方案而流行起来的。网络地址转换在很多国家广泛使用。所以NAT就成了家庭和小型办公室网络连接上的路由器的一个标准特征，因为对他们来说，申请独立的IP地址的代价要高于所带来的效益。
- 在一个典型的配置中，一个本地网络使用一个专有网络的指定子网（比如192.168.x.x或10.x.x.x）和连在这个网络上的一个路由器。这个路由器占有这个网络地址空间的一个专有地址（比如192.168.0.1），同时它还通过一个或多个因特网服务提供商提供的公有的IP地址（叫做“过载”NAT）连接到因特网上。当信息由本地网络向因特网传递时，源地址从专有地址转换为公用地址。由路由器跟踪每个连接上的基本数据，主要是目的地址和端口。当有回复

返回路由器时，它通过输出阶段记录的连接跟踪数据来决定该转发给内部网的哪个主机；如果有多个公用地址可用，当数据包返回时，TCP或UDP客户机的端口号可以用来分解数据包。对于因特网上的通信，路由器本身充当源和目的。

- NAPT扩大了2的16次方的ip地址数量（端口号的数量）

Windows NAT表

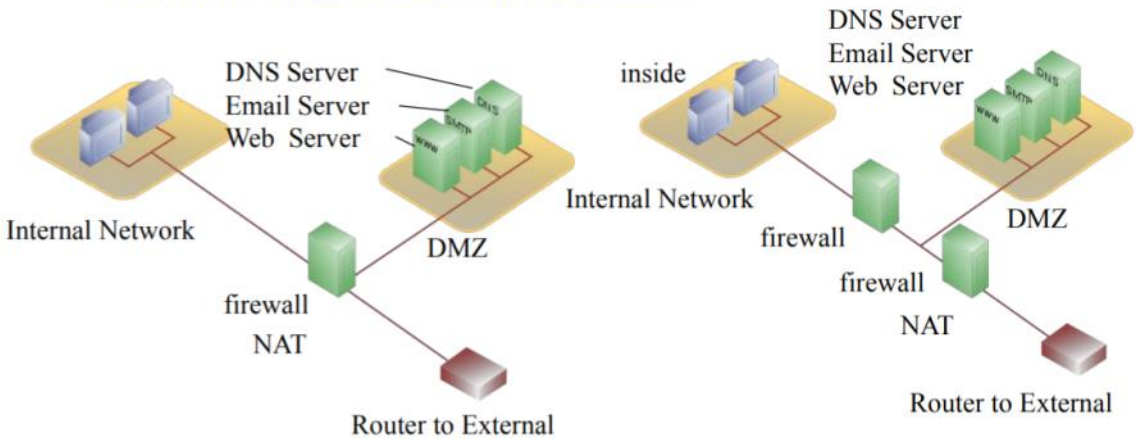
静态映射，没有ttl，一直保存

SIST-STU1 - 网络地址转换会话映射表格

通讯协议	方向	专用地址	专用端口	公用地址	公用端口	远程地址	远程端口	空闲时间
TCP	出站	10.1.100.9	2,130	172.18.186.211	62,447	210.28.176.12	80	8
TCP	出站	10.1.100.9	2,131	172.18.186.211	62,448	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,132	172.18.186.211	62,449	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,133	172.18.186.211	62,450	64.233.181.154	80	11
TCP	出站	10.1.100.9	2,134	172.18.186.211	62,451	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,135	172.18.186.211	62,452	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,136	172.18.186.211	62,453	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,137	172.18.186.211	62,454	64.233.181.154	80	11
TCP	出站	10.1.100.9	2,138	172.18.186.211	62,455	119.42.233.243	80	11
TCP	出站	10.1.100.9	2,139	172.18.186.211	62,456	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,140	172.18.186.211	62,457	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,141	172.18.186.211	62,458	210.28.176.12	80	11
TCP	出站	10.1.100.9	2,142	172.18.186.211	62,459	64.233.181.154	80	11
TCP	出站	10.1.100.9	2,143	172.18.186.211	62,460	64.233.181.154	80	11
TCP	出站	10.1.100.9	2,159	172.18.186.211	62,476	210.28.176.12	80	10
TCP	出站	10.1.100.9	2,160	172.18.186.211	62,477	210.28.176.12	80	10
TCP	出站	10.1.100.9	2,161	172.18.186.211	62,478	210.28.176.12	80	10
TCP	出站	10.1.100.9	2,162	172.18.186.211	62,479	210.28.176.12	80	10

非军事化区*

- 非军事化区(Demilitarized Zone, DMZ)是位于内部网络和外部网络之间并为双方提供因特网服务的区域。
- 内网主机可以访问内网主机、DMZ和因特网。内网主机可以使用内部地址或全局地址访问DMZ的服务器。外部主机只能通过全局地址访问DMZ的服务器，不能访问内网主机。



DMZ是英文“demilitarized zone”的缩写，中文名称为“隔离区”。它是为了解决安装防火墙后外部网络不能访问内部网络服务器的问题，而设立的一个非安全系统与安全系统之间的缓冲区，这个缓冲区位于企业内部网络和外部网络之间的小网络区域内，在这个小网络区域内可以放置一些必须公开的服务器设施，如企业Web服务器、FTP服务器和论坛等。另一方面，通过这样一个DMZ区域，更加有效地保护了内部网络，因为这种网络部署，比起一般的防火墙方案，对攻击者来说又多

了一道关卡。

多播IP地址*

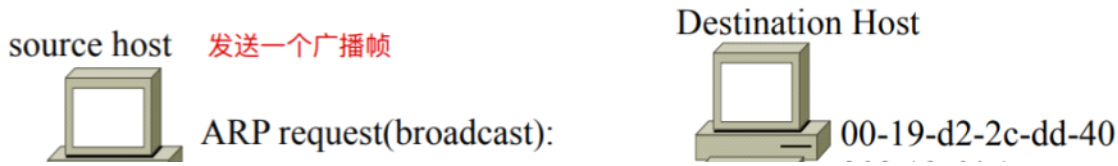
多播地址范围	用法
224.0.0.0~239.255.255.255	IPv4的多播地址空间
224.0.0.0~224.0.0.255	由IANA分配的永久地址。路由器不转发目的地址为这些地址的IP数据包
224.0.1.0~224.0.1.255	由IANA分配的永久地址。路由器会转发目的地址为这些地址的IP数据包
232.0.0.0~232.255.255.255	用于指定源的多播应用
233.0.0.0~233.255.255.255	由AS分配的全局多播地址
239.0.0.0~239.255.255.255	私有多播地址
其它地址	临时多播地址(transient address)

知名多播地址*

224.0.0.0	base address (reserved)		RFC 1222
224.0.0.1	All System on this subset	本网中的所有节点	RFC 1222
224.0.0.2	All routers on this subset	本网中的所有路由器	RFC 1222
224.0.0.3	Unassigned		
224.0.0.4	DVMRP		RFC 1075
224.0.0.5	OSPFIGP-all routers	所有OSPF路由器	RFC 1583
224.0.0.6	OSPFIGP-designated routers	所有OSPF指派路由器	RFC1583
224.0.0.7	ST routers		RFC 1190
224.0.0.8	ST hosts		RFC 1190
224.0.0.9	RIP2	所有RIPv2路由器	RFC 1723
224.0.0.10	IGRP routers	所有IGRP路由器	Cisco
224.0.0.11	Mobile-agents		
224.0.0.12	DHCP server/relay agent	所有DHCP路由器	RFC 1884
224.0.0.13	PIM	所有PIM路由器	RFC 1884
244.0.0.14-224.0.0.255	unassigned		

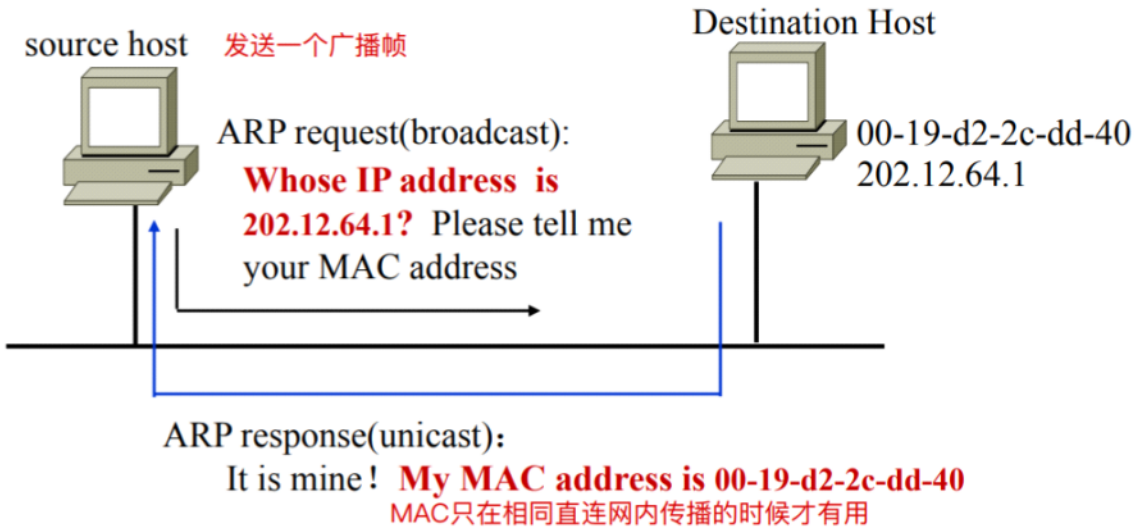
ARP协议 (1)

- 地址解析协议 (Address Resolution Protocol)可以把IP地址映射为MAC地址。



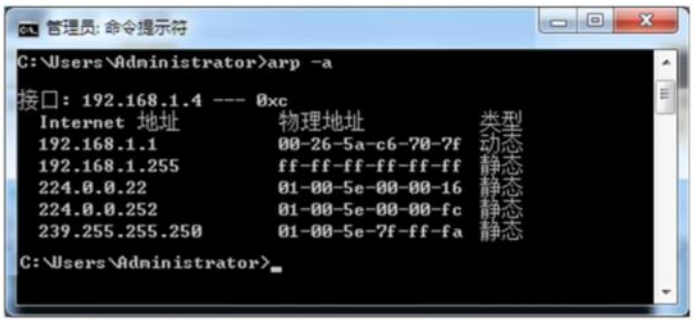
ARP协议 (1)

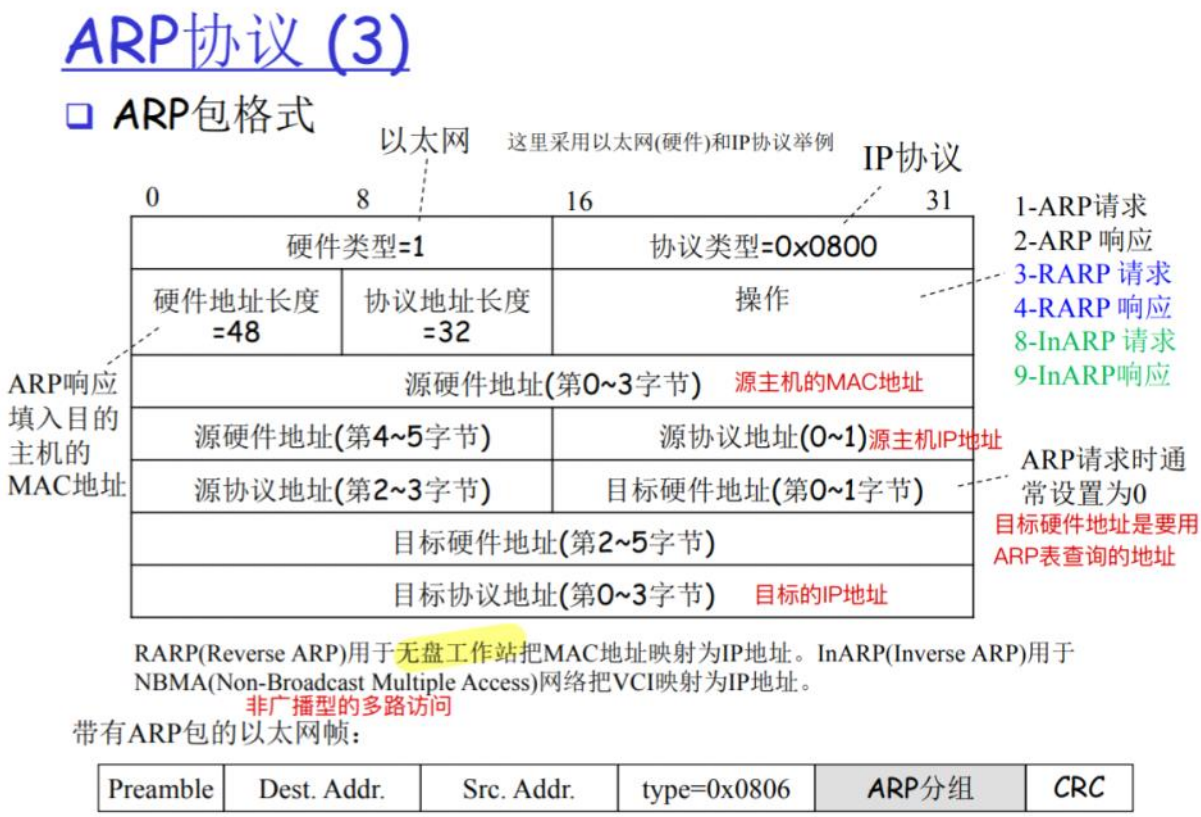
- 地址解析协议 (Address Resolution Protocol)可以把IP地址映射为MAC地址。



ARP协议 (2)

- ARP协议没有超时重传机制。超时没有收到响应，则丢弃引发ARP查询的IP分组。
- 源主机获得的映射结果缓存在ARP表中<IP address，MAC address，TTL>。TTL超时则会删除对应的ARP表项，TTL取值由系统确定，一般为2~20分钟。
- 当收到ARP请求，目的主机会缓存源主机的映射，其它主机如果已缓存该映射，则会重置TTL。
- 也可以把映射直接加入ARP缓存，称为静态ARP映射。静态ARP映射不会因超时而被删除。
- Windows中的ARP表：





ARP协议工作机制：

目标以太网地址：目标MAC地址。FF:FF:FF:FF:FF:FF （二进制全1）为广播地址。

源以太网地址：发送方MAC地址。

帧类型：以太类型，ARP为0x0806。

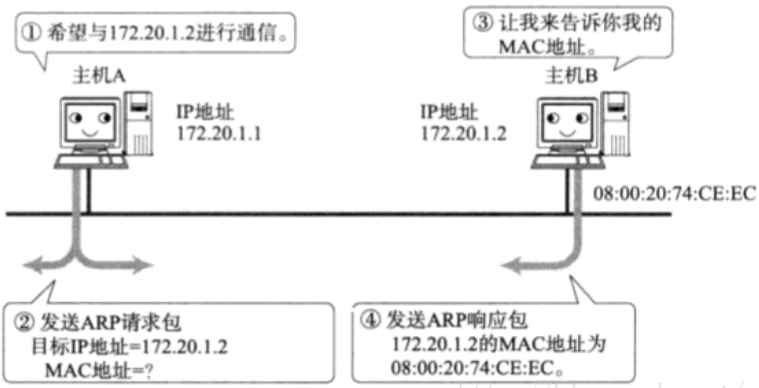
在每台安装有TCP/IP协议的电脑或路由器里都有一个ARP缓存表，表里的IP地址与MAC地址是一对应的，如下表所示。

主机名称	IP地址	MAC地址
A	192.168.38.10	00-AA-00-62-D2-02
B	192.168.38.11	00-BB-00-62-C2-02
C	192.168.38.12	00-CC-00-62-C2-02
D	192.168.38.13	00-DD-00-62-C2-02
E	192.168.38.14	00-EE-00-62-C2-02
...	...	...

当主机A要向本局域网上的某个主机B发送IP数据报时，就先在其ARP高速缓存中查找有无主机B的IP地址。如果有，就在ARP高速缓存中查出其对应的硬件地址，再把其硬件地址写入到MAC帧，然后通过局域网把该MAC帧发往此硬件地址。

如果主机高速缓存中没有则运行ARP按照以下步骤查找出主机B的硬件地址。

（1）ARP进程在本局域网上广播发送一个ARP请求分组如下：



（2）本局域网上所有的主机上运行的ARP进程都收到此ARP请求分组。

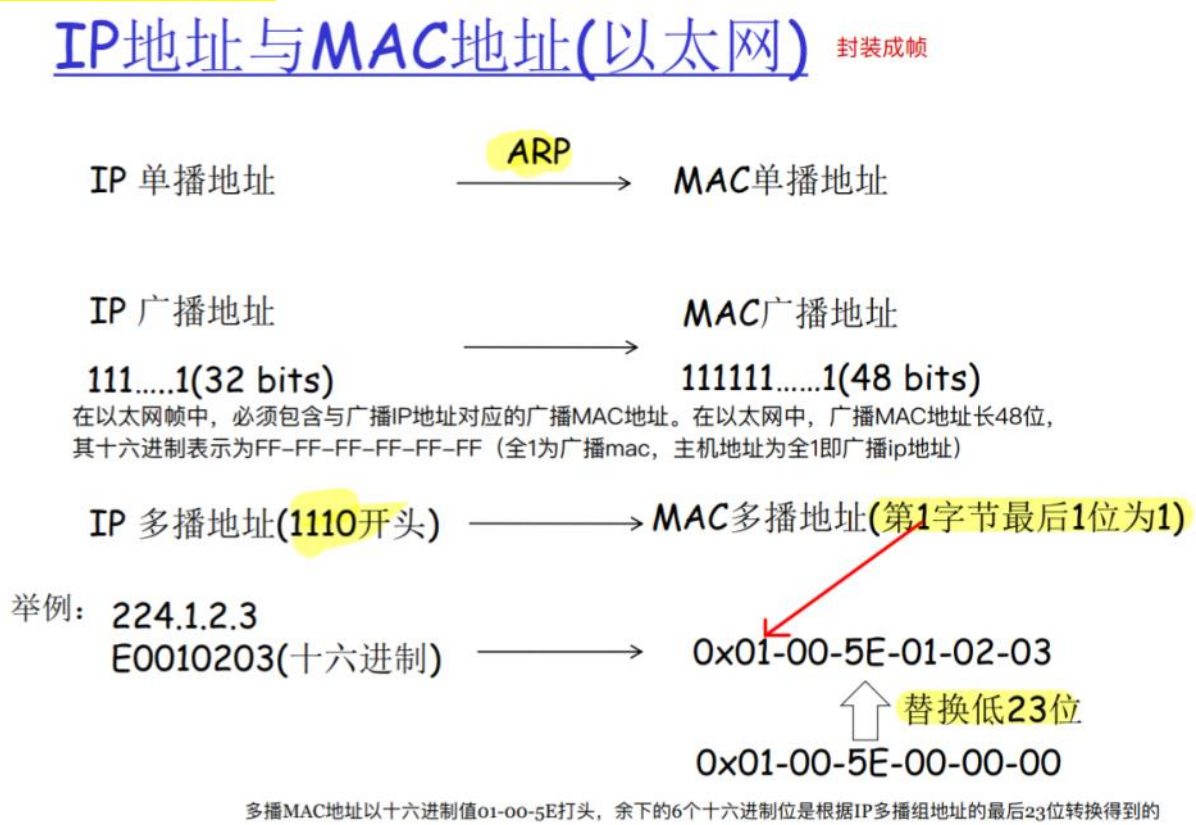
（3）主机B在ARP分组中见到自己的IP地址就向A发送ARP响应分组，并写入自己的硬件地址，相应分组是普通的单播。

（4）主机A收到主机B的ARP响应分组后，就在其ARP高速缓存中写入主机B的IP地址到硬件地址的映射。

（5）另外，当发送主机和目的主机不在同一个局域网中时，即便知道目的主机的MAC地址，两者也不能直接通信，必须经过路由转发才可以。所以此时，发送主机通过ARP协议获得的将不是目的主机的真实MAC地址，而是一台可以通往局域网外的路由器的MAC地址。于是此后发送主机发往目的主机的所有帧，都将发往该路由器，通过它向外发送。这种情况称为委托ARP或ARP代理（ARP Proxy）。

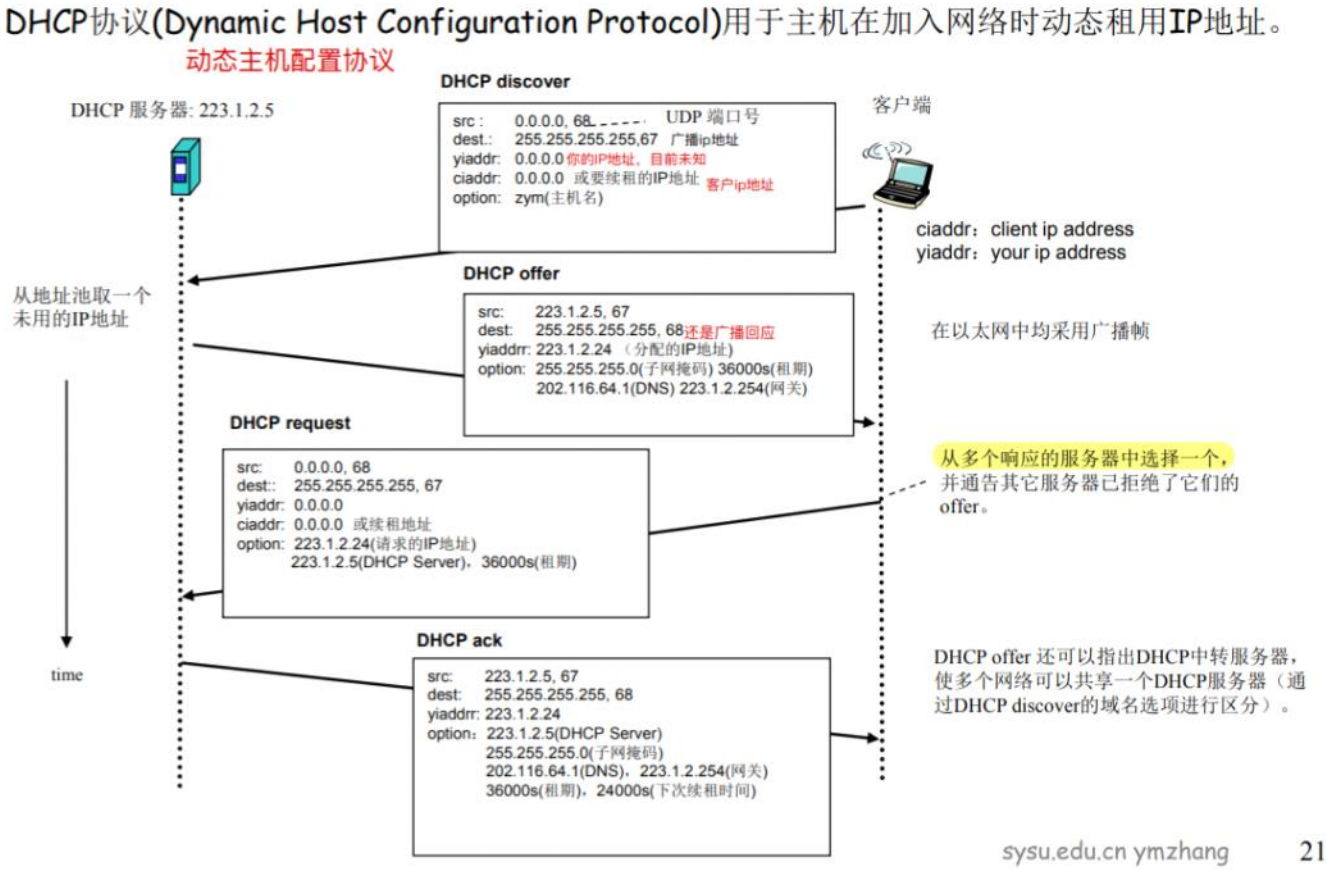
IP协议就是0x0800

ARP协议就是0x0806



DHCP协议

http://www.ietf.org/rfc/rfc2131.txt



在IP地址的动态分配中，DHCP客户端想DHCP服务器发送IP地址请求。DHCP服务器会维护一个 IP 地址池，DHCP从地址池从取出一个IP回应给DHCP客户端。在地址分配时，DHCP服务器也会指定回应给DHCP客户端的IP地址的租约期，该地址只有在该租约期内可用，不过DHCP客户端可用在租约期内请求延长租约（更新租约期）

DHCP 协议 操作流程

1. 发现阶段，即DHCP客户机寻找DHCP服务器的阶段。DHCP客户机以广播方式（因为DHCP服务器的IP地址对于客户机来说是未知的）发送DHCP discover发现信息来寻找DHCP服务器，即向地址255.255.255.255发送特定的广播信息。网络上每一台安装了TCP/IP协议的主机都会接收到这种广播信息，但只有DHCP服务器才会做出响应。

2. 提供阶段，即DHCP服务器提供IP地址的阶段。在网络中接收到DHCP discover发现信息的DHCP服务器都会做出响应，它从尚未出租的IP地址中挑选一个分配给DHCP客户机，向DHCP客户机发送一个包含出租的IP地址和其他设置的DHCP offer提供信息。

3. 选择阶段，即DHCP客户机选择某台DHCP服务器提供的IP地址的阶段。如果有多台DHCP服务器向DHCP客户机发来的DHCP offer提供信息，则DHCP客户机只接受第一个收到的DHCP offer提供信息，然后它就以广播方式回答一个DHCP request请求信息，该信息中包含向它所选定的DHCP服务器请求IP地址的内容。之所以要以广播方式回答，是为了通知所有的DHCP服务器，他将选择某台DHCP服务器所提供的IP地址。

4. 确认阶段，即DHCP服务器确认所提供的IP地址的阶段。当DHCP服务器收到DHCP客户机回答的DHCP request请求信息之后，它便向DHCP客户机发送一个包含它所提供的IP地址和其他设置的DHCP ack确认信息，告诉DHCP客户机可以使用它所提供的IP地址。然后DHCP客户机便将其TCP/IP协议与网卡绑定，另外，除DHCP客户机选中的服务器外，其他的DHCP服务器都将收回曾提供的IP地址。

5. 重新登录。以后DHCP客户机每次重新登录网络时，就不需要再发送DHCP discover发现信息了，而是直接发送包含前一次所分配的IP地址的DHCP request请求信息。当DHCP服务器收到这一信息后，它会尝试让DHCP客户机继续使用原来的IP地址，并回答一个DHCP ack确认信息。如果此IP地址已无法再分配给原来的DHCP客户机使用时（比如此IP地址已分配给其它DHCP客户机使用），则DHCP服务器给DHCP客户机回答一个DHCP nack否认信息。当原来的DHCP客户机收到此DHCP nack否认信息后，它就必须重新发送DHCP discover发现信息来请求新的IP地址。

6. 更新租约。DHCP服务器向DHCP客户机出租的IP地址一般都有一个租借期限，期满后DHCP服务器便会收回出租的IP地址。如果DHCP客户机要延长其IP租约，则必须更新其IP租约。DHCP客户机启动时和IP租约期限过一半时，DHCP客户机都会自动向DHCP服务器发送更新其IP租约的信息。

ciaddr只有在续租的时候或者不需要discover时才不是0.0.0.0

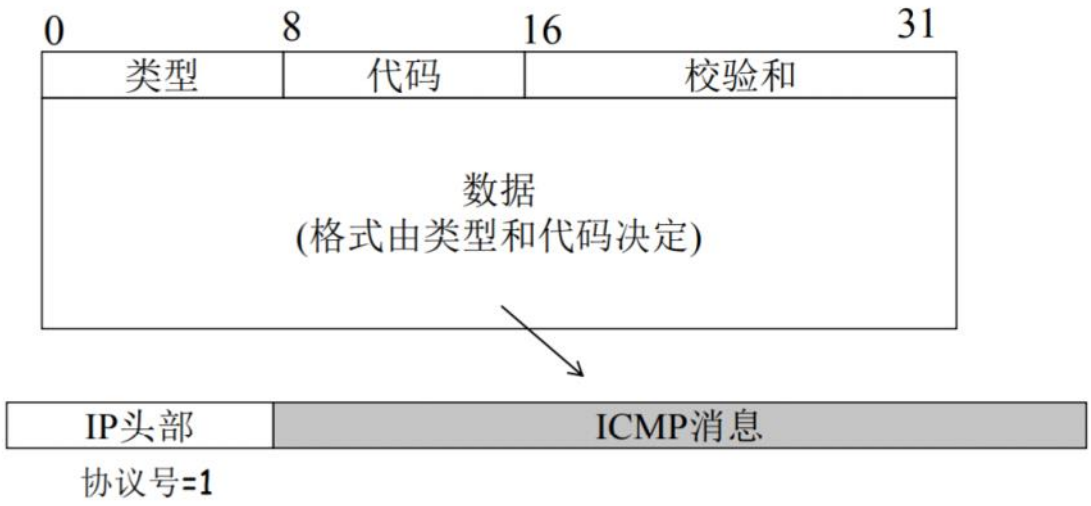
ICMP协议

- 因特网控制消息协议(Internet Control Message Protocol)用于主机或路由器发布网络级别的控制消息。 http://tools.ietf.org/html/rfc792
- ICMP消息的常见类型：

类型	代码	描述	查询	差错
0	0	回显应答(Ping应答)	√	
8	0	请求回显(Ping请求)	√	
3		目标不可达		
	0	网络不可达		√
	1	主机不可达		√
	2	协议不可达		√
	3	端口不可达		√
	4	需要分段但不可分段(DF=1)		√
	5	源站选路失败(IP Options)		√
	11	由于TOS, 网络不可达		√
4	0	源端抑制 (控制源主机发送速度)		√

类型	代码	描述	查询	差错
5		重定向		
	1	对主机重定向		√
11		超时		
	0	传输期间TTL=0		√
	1	数据报重组超时		√
12		参数问题		
	0	坏的IP头部(各种错误)		√
	1	缺少必要的选项		√
13	0	时间戳请求	√	
14	0	时间戳应答	√	

- ICMP消息的一般格式



ICMP不可达消息

8b	8b	16b
类型=3	代码=0~15	校验和
未用(必须填为0)		
原IP头部+原IP数据部份的头64比特		

代码	含义	说明
0	网络不可达	目的地址为私有地址，路由表出错等
1	主机不可达	不能找到到目的网络的路由 查路由表，没有ip地址就给源主机传
2	协议不可达	上层协议不存在
3	端口不可达	只用于UDP协议，UDP端口号没有绑定进程。对于TCP协议，如果出现没有进程绑定端口，则会发送TCP 复位消息而不是本消息
4	分段错误	需要分段但是设置了DF
5	源路由错误	IP源路由选项出错

ICMP回响请求和答复消息(Ping)

8b	8b	16b
类型=0或8	代码=0	检验和
标识符		序号
数据		

回响(echo)请求：类型 = 8; 回响答复：类型= 0

- 从回响请求收到的标识符、序号和数据将拷贝到回响答复中。
- 标识符和序号用来区分不同的响应。例如：把进程号记录为标识符，并记录发送序号。

ICMP时间超时消息

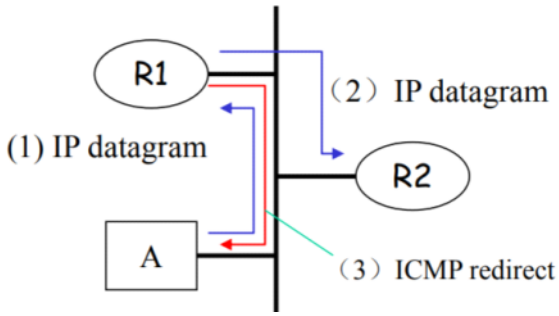
8b	8b	16b
类型=11	代码=0或1	校验和
未用(必须为0)		
原IP头部+原IP数据部份的头64比特		

代码	说明
0	TTL减为0
1	重组IP片段超时

ICMP重定位消息

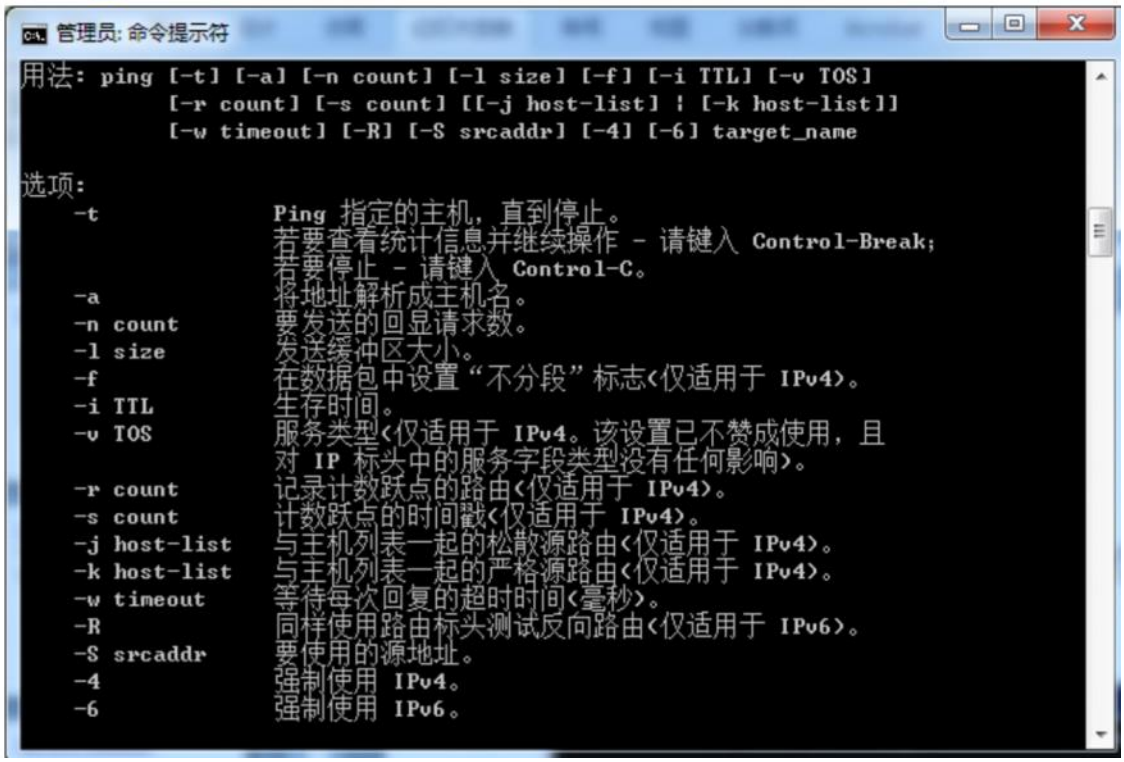
8b	8b	16b
类型=5	代码	校验和
未用(必须为0)		
原IP头部+原IP数据部份的头64比特		

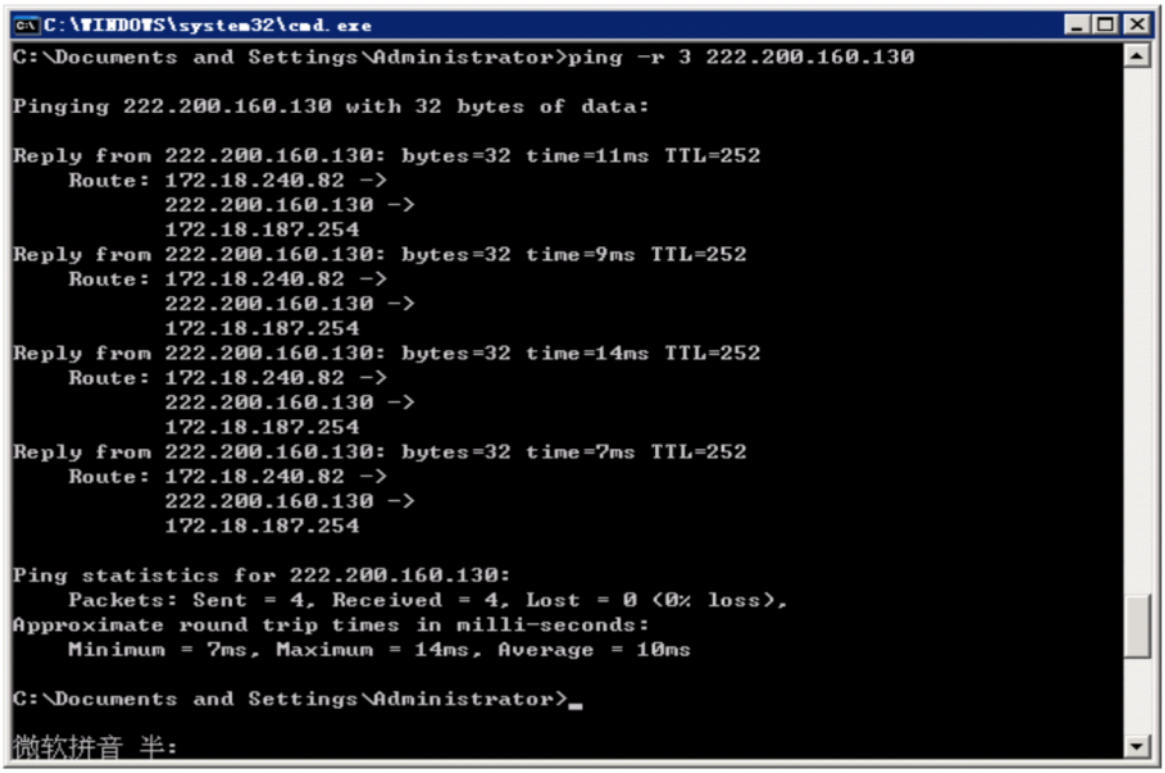
Code	Description
0	Redirect datagrams for the Network.
1	Redirect datagrams for the Host.
2	Redirect datagrams for the Type of Service and Network.
3	Redirect datagrams for the Type of Service and Host.



当R1发现把一个数据包转发给R2的接口就是其接收接口，则会把从网络重定向消息发给主机A，要主机A直接把发往这些网络的数据报直接发给R2。

Ping 和 ICMP协议





ping 原理

ping 程序是用来探测主机到主机之间是否可通信，如果不能ping到某台主机，表明不能和这台主机建立连接。ping 使用的是ICMP协议，它发送icmp回送请求消息给目的主机。ICMP协议规定：目的主机必须返回ICMP回送应答消息给源主机。如果源主机在一定时间内收到应答，则认为主机可达。（见上图类型为0或8）

首先，Ping命令会构建一个固定格式的ICMP请求数据包，然后由ICMP协议将这个数据包连同地址

“192.168.1.2”一起交给IP层协议（和ICMP一样，实际上是一组后台运行的进程），IP层协议将以地址“192.168.1.2”作为目的地址，本机IP地址作为源地址，加上一些其他的控制信息，构建一个IP数据包，并在一个映射表中查找出IP地址192.168.1.2所对应的物理地址（也叫MAC地址，熟悉网卡配置的朋友不会陌生，这是数据链路层协议构建数据链路层的传输单元——帧所必需的），一并交给数据链路层。后者构建一个数据帧，目的地址是IP层传过来的物理地址，源地址则是本机的物理地址，还要附加上一些控制信息，依据以太网的介质访问规则，将它们传送出去。
其中映射表由ARP实现。ARP(Address Resolution Protocol)是地址解析协议,是一种将IP地址转化成物理地址的协议。ARP具体说来就是将网络层（IP层，也就是相当于OSI的第三层）地址解析为数据连接层（MAC层，也就是相当于OSI的第二层）的MAC地址。

主机B收到这个数据帧后，先检查它的目的地址，并和本机的物理地址对比，如符合，则接收；否则丢弃。接收后检查该数据帧，将IP数据包从帧中提取出来，交给本机的IP层协议。同样，IP层检查后，将有用的信息提取后交给ICMP协议，后者处理后，马上构建一个ICMP应答包，发送给主机A，其过程和主机A发送ICMP请求包到主机B一模一样。