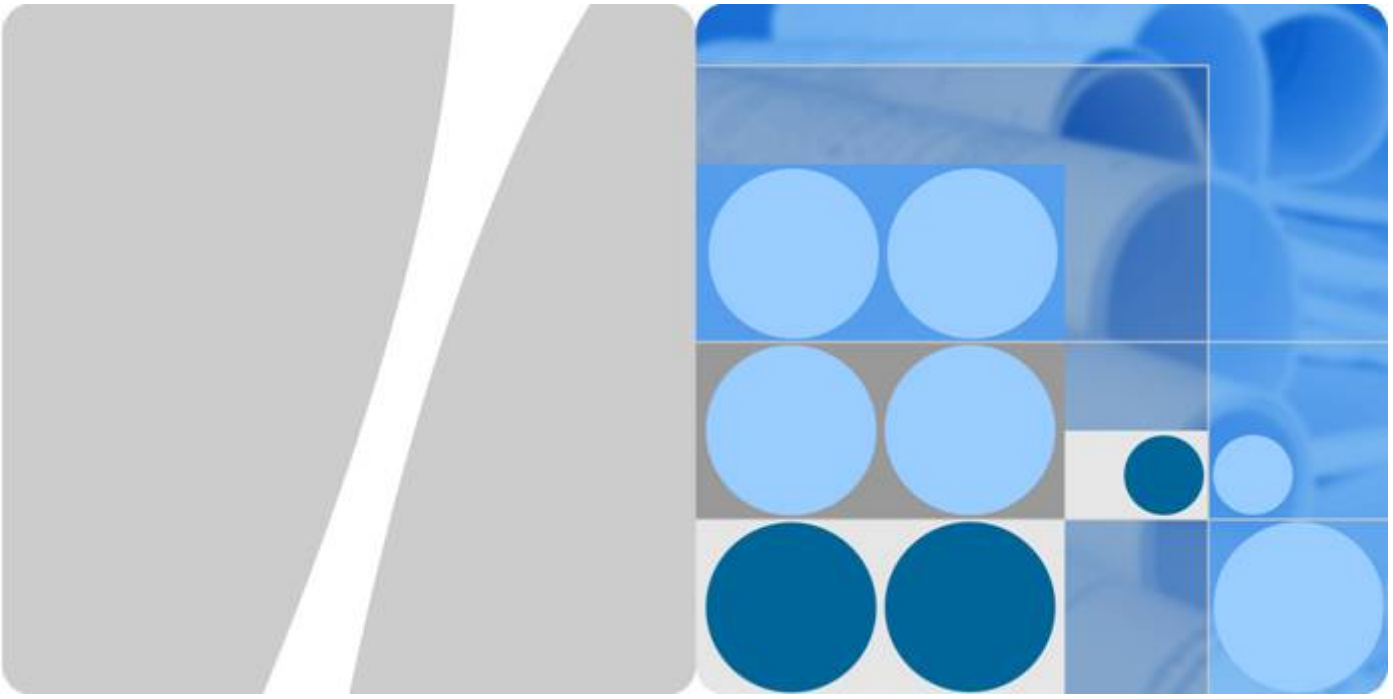


# Huawei AR200-S 系列企业路由器

# V200R002C00

# 特性描述-IP 业务

文档版本　01

发布日期　2011-12-30

华为技术有限公司

## 商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

## 注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

# 华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

# 前 言

## 读者对象

本文档针对 IP 业务特性，从简介、原理描述和应用三个方面介绍了 IP 业务特性。

本文档与其它类型手册相结合，便于读者深入掌握特性的实现原理。

本文档主要适用于以下工程师：

- 网络规划工程师
- 调测工程师
- 数据配置工程师
- 系统维护工程师

## 符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

| 符号 | 说明 |
| --- | --- |
| 危险 | 以本标志开始的文本表示有高度潜在危险，如果不能避免，会导致人员死亡或严重伤害。 |
| 警告 | 以本标志开始的文本表示有中度或低度潜在危险，如果不能避免，可能导致人员轻微或中等伤害。 |
| 注意 | 以本标志开始的文本表示有潜在风险，如果忽视这些文本，可能导致设备损坏、数据丢失、设备性能降低或不可预知的结果。 |
| 窍门 | 以本标志开始的文本能帮助您解决某个问题或节省您的时间。 |
| 说明 | 以本标志开始的文本是正文的附加信息，是对正文的强调和补充。 |

# 命令行格式约定

| 格式 | 意义 |
| --- | --- |
| **粗体** | 命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。 |
| *斜体* | 命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。 |
| [ ] | 表示用“[ ]”括起来的部分在命令配置时是可选的。 |
| { x \| y \| ... } | 表示从两个或多个选项中选取一个。 |
| [ x \| y \| ... ] | 表示从两个或多个选项中选取一个或者不选。 |
| { x \| y \| ... }* | 表示从两个或多个选项中选取多个，最少选取一个，最多选取所有选项。 |
| [ x \| y \| ... ]* | 表示从两个或多个选项中选取多个或者不选。 |
| &<1-n> | 表示符号&前面的参数可以重复 1 ～ n 次。 |
| # | 由“#”开始的行表示为注释行。 |

# 修订记录

修改记录累积了每次文档更新的说明。最新版本的文档包含以前所有文档版本的更新内容。

## 文档版本 01 (2011-12-30)

第一次正式发布。

# 目 录

# 1 IP 地址

## 关于本章

# 1.1 介绍

在 IP 网络上，需要为网络上的主机分配 IP 地址。如果用户要将一台计算机连接到 Internet 上，就需要向 ISP 申请一个 IP 地址。

IP 地址是在计算机网络中被用来唯一标识一台设备的一组数字，各个节点（设备）之间使用 IP 协议进行通信。IP 地址的层次是按网络结构进行划分，一个 IP 地址是由网络号和主机号两部分组成。

IP 地址由 32 位二进制数值组成，但为了便于用户识别和记忆，采用了“点分十进制表示法”。采用了这种表示法的 IP 地址由 4 个点分十进制整数来表示，每个十进制整数对应一个字节。例如，A 主机的 IP 地址使用二进制的表示形式为 00001010 00000001 00000001 00000010，采用点分十进制表示法表示为 10.1.1.2。

IP 地址由如下两部分组成：

- 网络号码字段（net-id）：用于区分不同网络。网络号码字段的前几位称为类别字段（又称为类别比特），用来区分 IP 地址的类型。
- 主机号码字段（host-id）：用于区分一个网络内的不同主机。

IP 地址的网络号码字段用来标识一个网络，主机号码字段用来标识网络中网络设备的一个连接。如果有多台网络设备，无论它们分别处于任何物理位置，只要它们具有相同的网络号，那他们就处在同一网络中。也就是说，在公共网络内的多台网络设备是否处于相同网络与它们所处的物理位置无关。

# 1.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC 1166 | Internet Numbers | - |
| RFC 1918 | Address Allocation for Private Internets | - |

# 1.3 可获得性

## 涉及网元

无需其它网元的配合。

## License 支持

无需获得 License 许可，均可获得该特性的服务。

### 版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

### 特性依赖

不依赖其他特性。

### 硬件要求

对硬件无特殊要求。

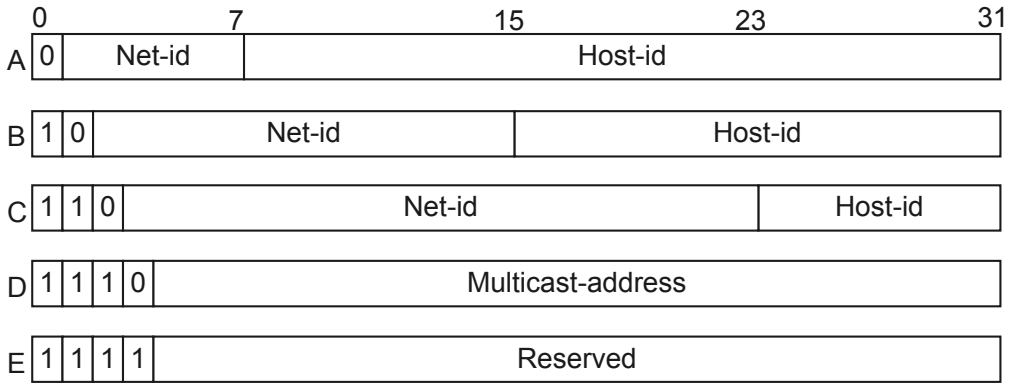
## 1.4 原理描述

### 1.4.1 IP 地址分类

为了方便 IP 地址的管理及组网，IP 地址分成五类，如图 1-1 所示。

通过网络号码字段的前几个比特就可以判断 IP 地址属于哪一类，这是区分各类地址最简单的方法。

**图 1-1** 五类 IP 地址




目前大量使用中的 IP 地址属于 A、B、C 三类 IP 地址中的一种。D 类地址是组播地址，E 类地址保留。在 IETF（Internet Engineering Task Force）发布的 RFC1166 Internet Numbers 中详细描述了各类 IP 地址。

在使用 IP 地址时要注意，一些 IP 地址是保留作为特殊用途的，一般的用户不能使用。表 1-1 列出各类 IP 地址的范围。

**表 1-1** IP 地址分类及范围

| 网络类型 | 地址范围 | 用户可用的IP 网络范围 | 说明 |
| --- | --- | --- | --- |
| A | 0.0.0.0 ～ 127.255.255.255 | 1.0.0.0 ～ 126.0.0.0 | 全 0 的主机号码表示该 IP 地址就是网络的地址，用于网络路由；全 1 的主机号码表示广播地址，即对该网络上所有的主机进行广播；IP 地址 0.0.0.0 仅在采用 DHCP 方式的系统启动时允许本主机利用它进行临时的通信，并且永远不是有效目的地址；网络号码为 0 的 IP 地址表示当前网络的主机，可以让机器引用自己的网络而不必知道其网络号；所有形如 127.X.Y.Z 的地址都保留作环回测试，发送到这个地址的分组不会输出到线路上，它们被内部处理并当作输入分组。 |
| B | 128.0.0.0 ～ 191.255.255.255 | 128.1.0.0 ～ 191.254.0.0 | 全 0 的主机号码表示该 IP 地址就是网络的地址，用于网络路由；全 1 的主机号码表示广播地址，即对该网络上所有的主机进行广播。 |
| C | 192.0.0.0 ～ 223.255.255.255 | 192.0.1.0 ～ 223.255.254.0 | 全 0 的主机号码表示该 IP 地址就是网络的地址，用于网络路由；全 1 的主机号码表示广播地址，即对该网络上所有的主机进行广播。 |
| D | 224.0.0.0 ～ 239.255.255.255 | 无 | D 类地址是一种组播地址。 |
| E | 240.0.0.0 ～ 255.255.255.255 | 无 | 保留。255.255.255.255 用于局域网广播地址。 |

## 1.4.2 IP 地址的特点

IP 地址的主要特点有：

- IP 地址是一种非等级的地址结构，不同于电话号码的结构。也就是说，IP 地址不能反映任何有关主机位置的地理信息，只能通过网络号码字段判断出主机属于哪个网络。
- 当一个主机同时连接到两个网络上时，该主机就必须同时具有两个相应的 IP 地址，其网络号码 net-id 是不同的，这种主机称为多地址主机（Multihomed Host）。主机上的每个接口都对应着一个 IP 地址，因此多接口主机会有多个 IP 地址。
- 按照 Internet 的观点，用转发器或网桥连接起来的若干个局域网仍为一个网络，因此这些局域网都具有同样的网络号码 net-id。
- 在 IP 地址中，所有分配到网络号码 net-id 的网络（不管是小的局域网还是很大的广域网）都是平等的。

## 1.4.3 特殊 IP 地址

在实际使用过程中，有一些特殊的 IP 地址，其范围和描述如表 1-2 所示。

**表 1-2** 特殊情况的 IP 地址

| IP 地址网络号 | IP 地址子网号 | IP 地址主机号 | 能否作为源端地址 | 能否作为目的端地址 | 描述 |
|---|---|---|---|---|---|
| 全 0 | - | 全 0 | 可以 | 不可以 | 用于网络上的主机 |
| 全 0 | - | 主机号 | 可以 | 不可以 | 用于网络上的特定主机 |
| 127 | - | 任何值 | 可以 | 可以 | 用于环回地址 |
| 全 1 | - | 全 1 | 不可以 | 可以 | 用于受限的广播（永远不被转发） |
| net-id | - | 全 1 | 不可以 | 可以 | 用于向以 net-id 为目的的网络广播 |
| net-id | subnet-id | 全 1 | 不可以 | 可以 | 用于向以 net-id，subnet-id 为目的的子网广播 |
| net-id | 全 1 | 全 1 | 不可以 | 可以 | 用于向以 net-id 为目的的所有子网广播 |

说明

net-id，subnet-id 分别表示不全为 0 和不全为 1 的对应字段。

### 1.4.4 私有 IP 地址

为了解决 IP 地址短缺的问题，提出了私有地址的概念。私有地址是指内部网络或主机地址，这些地址只能用于某个内部网络，不能用于公共网络。RFC1918 描述了为私有网络预留的 3 个 IP 地址段。

IP 地址分配组织规定将下列的 IP 地址保留用作私有地址，如表 1-3 所示。

**表 1-3** 私有 IP 地址

| 网络类型 | 地址范围 |
|---|---|
| A | 10.0.0.0 ～ 10.255.255.255 |
| B | 172.16.0.0 ～ 172.31.255.255 |
| C | 192.168.0.0 ～ 192.168.255.255 |

## 1.5 应用

### 1.5.1 子网划分

IP 地址的网络部分称为网络地址，网络地址用于唯一的标识一个网段。通过将网络地址进一步划分为若干个子网，实现不同子网之间隔离广播报文。

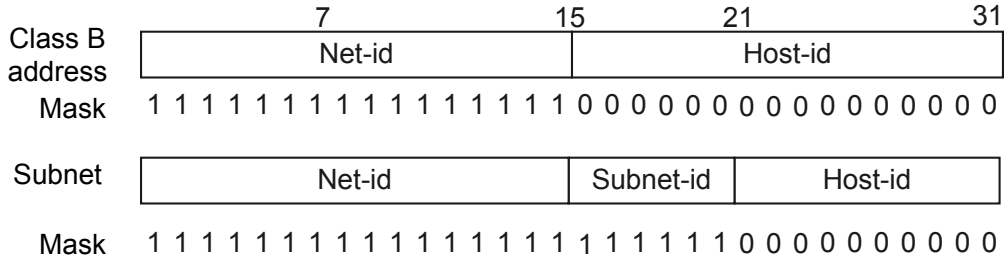
从地址分配的角度来看，子网是网段地址的扩充。为了使 IP 地址的使用更加灵活，只分配 IP 地址的网络号码 net-id，而后面的主机号码 host-id 则是受本单位控制。即某个单位申请到 IP地址时，实际上只是拥有了一个网络号码 net-id，具体的各个主机号码 host-id 则由该单位自行分配，只要做到在该单位管辖的范围内无重复的主机号码即可。

当一个单位的主机很多而且分布在很大的地理范围时，为了便于管理，可将单位内部的主机号码再进一步划分为多个子网。通过子网划分，整个网络地址可以划分成更多的小网络。

子网的划分是网络内部的行为，从外部看，这个单位只有一个网络号码。只有当外部的报文进入到本单位范围后，本单位的路由设备才根据子网号码再进行选路，找到目的主机。

如图 1-2 所示，为一个 B 类 IP 地址子网划分情况，其中子网掩码由一串连续的“1”和一串连续的“0”组成。“1”对应于网络号码和子网号码字段，而“0”对应于主机号码字段。

**图 1-2** IP 地址子网划分




将 32 位的 IP 地址和子网掩码的对应位作与运算可以确定 IP 地址的网络号。例如，IP 地址为 10.1.1.2，子网掩码为 255.255.0.0，那么将 IP 地址与其相应掩码位执行与运算的结果就是网络地址 10.1.0.0。

多划分出一个子网号码字段是要付出代价的。举例来说，本来一个 B 类 IP 地址可以容纳 65534 个主机号码。但划分出 6bits 长的子网字段后，最多可有 64 个子网，每个子网有 10bit 的主机号码，即每个子网最多可有 1022（$2^{10}$-2，去掉全 1 和全 0 的主机号码）个主机号码。因此主机号码的总数是（64 x 1022 = 65408）个，比不划分子网时要少 126 个。

若一个单位不进行子网的划分，则其子网掩码即为默认值，此时子网掩码中“1”的长度就是网络号码的长度。因此，对于 A、B 和 C 类的 IP 地址，其对应子网掩码的默认值分别为 255.0.0.0、255.255.0.0.和 255.255.255.0。

子网划分与 IP 地址规划时，通常需要综合考虑以下原则，实现合理高效的网络规划。

## 层次性

实现网络的层次性划分，需要综合考虑地域和业务因素，尽可能和网络层次相对应，采用自顶向下的方法划分,达到有效管理网络、简化路由表的目标。一般情况下：

- 对于大骨干网络和大城域网络相结合的网络，采用扁平化思路划分方式。
- 对于行政区类型的网络，采用多级网络分配方式。

### 连续性

连续地址在层次结构的网络中易于进行路由聚合，大大缩减路由表数量，提高路由查找的效率。

- 尽量为每个区域分配连续的 IP 地址空间。
- 尽量为具有相同业务和功能的设备分配连续的 IP 地址。
- 即使使用了支持地址重叠的 MPLS/VPN 技术，也尽量不要规划为相同的地址。

### 扩展性

分配地址时，在每一层次上都要留有余量。当网络规模扩展时能保证地址分配的连续性，实现网络的长远规划。

骨干网络应有足够的连续地址组成独立的自治域，并为今后的扩展留有余地。

### 高效性

划分子网时，要保证充分利用地址资源，使子网的划分满足主机个数的要求。

- 利用可变长子网掩码（VLSM）技术，分配 IP 地址，充分合理地利用地址资源。
- 与网络的路由机制设计相结合，合理使用已划分的地址空间，提高地址的利用率。

### 业务相关性

规划 IP 地址时，应该为具有类似功能的设备分配相同类型的 IP 地址。

- 对于高端路由器、IP 电话网关、IP 电话网守、各种 Internet 服务器、防火墙、边缘或接入路由器等设备，应分配公网 IP 地址。VPN 与采用 VPN 方式进行的服务可以在 VPN 内部分配私网地址。
- 对于作为设备管理地址的 Loopback 接口，应尽量为其分配单独的一段连续地址，掩码使用 32 位。
- 对于设备间的互联接口，应尽量为其分配单独的一段连续地址，掩码使用 30 位。

## 1.5.2 IP 地址分配

用户访问 Internet 必须要有合法的 IP 地址，因此，用户地址的统一分配和管理是宽带接入服务器必须具备的基本功能。目前有以下几种主要的 IP 地址分配方式。

### 手工分配 IP 地址

可以直接在用户计算机上手工配置 IP 地址，这种方式一般用于固定用途的服务器或有特殊需要的用户，例如 Web 服务器、路由器等。为防止这类 IP 地址被盗用，可以在宽带接入服务器上配置 IP/VLAN、IP/PVC 绑定。

### 为 PPP 接入的用户分配 IP 地址

采用 PPP 方式接入的用户，可以利用 PPP 的地址协商功能，由接入服务器分配 IP 地址。有两种方法可以为 PPP 用户分配 IP 地址。

- 多用户情况下，配置 IP 地址池，在接口视图下指定该接口使用的地址池。
- 单用户情况下，不配置 IP 地址池，在接口上直接给用户配置指定的 IP 地址。

### 使用 DHCP 服务分配 IP 地址

DHCP 采用客户/服务器通信模式。网络管理员在 DHCP 服务器上设定一个 IP 地址范围，客户端向服务器提出配置申请（包括分配的 IP 地址、子网掩码、缺省网关等参数），服务器根据策略返回相应的配置信息。

采用以太网接入的用户，如 IPoEoA（IP over Ethernet over AAL5）、VLAN 用户，可以通过 DHCP 获得 IP 地址。

## 1.5.3 IP 地址借用

一个接口如果没有 IP 地址就无法生成路由，也就无法转发报文。IP 地址借用（IP Address Unnumbered）就是在本接口没有 IP 地址的情况下，可以使用其它接口的 IP 地址。所谓“借用 IP 地址”，其实质就是：一个接口上没有配置 IP 地址，但是还想使用该接口,就向其它有 IP 地址的接口借一个 IP 地址过来，以使该接口能够正常使用。

IP 地址借用的主要目的是节省 IP 地址资源。有时某个接口只是偶尔使用，这种情况也可配置该接口借用其他接口的 IP 地址，而不必让其一直占用一个单独的 IP 地址。

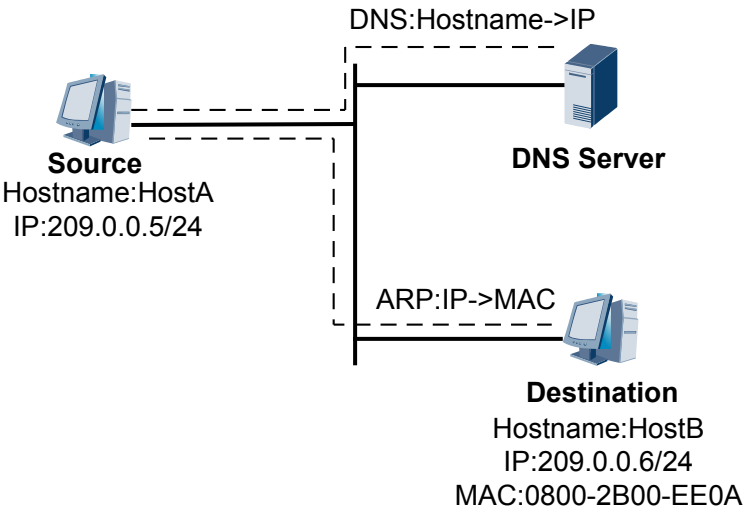
## 1.5.4 IP 地址解析

一台路由设备用来连接多个网络，具有多个网络的 IP 地址。上面讲的 IP 地址还不能直接用来进行通信。具体原因如下：

- IP 地址只是主机在网络层中的地址，若要将网络层中传送的数据报交给目的主机，必须知道该主机的物理地址。因此必须将 IP 地址解析为物理地址。
- 用户平时不愿意使用难于记忆的 IP 地址，而更愿意使用易于记忆的主机名，因此也需要将主机名解析为 IP 地址。

图 1-3 表示了主机名、IP 地址和物理地址之间的关系。在 Ethernet 上，主机的物理地址就是指 MAC 地址。将主机名解析为 IP 地址的操作是由 DNS 服务器来完成，而将 IP 地址解析为 MAC 地址的操作是由 ARP 来完成的。

**图 1-3 主机名、IP 地址和物理地址之间的关系**



### 1.5.5 广域网接口 IP 地址与链路层协议地址的映射

在路由设备中，除了维护以太网口 IP 地址到 MAC 地址的映射外，还需维护广域网口的 IP 地址与链路层协议地址的映射，这类映射有：

- 在封装帧中继接口上，IP 地址与 DLCI（Data Link Control Identifier）的映射。
- 在 ATM 接口上，IP 地址与 PVC 的映射。

上述这些映射，又可称为二次路由。为了保证 IP 业务能够运行在这些广域网链路上，需要正确配置这些地址映射。

## 1.6 术语与缩略语

### 术语

| 术语 | 解释 |
| --- | --- |
| 点分十进制表示法 | 点分十进制表示法是一种书写格式。采用了点分十进制的 IP 地址，即 IP 地址被“.”分隔成四部分，每部分都由十进制数字来表示。 |
| IP 地址借用 | 在本接口没有 IP 地址的情况下，使用其它接口的 IP 地址。 |
| 私有 IP 地址 | 指内部网络或主机地址，这些地址只能用于某个内部网络，不能用于公共网络。 |
| 子网掩码 | 子网掩码是 32 比特的二进制数字，使用子网掩码可以了解 IP 地址的网络号。 |

# 2 ARP

## 关于本章

# 2.1 介绍

## 定义

ARP（Address Resolution Protocol）是用来将 IP 地址解析为 MAC 地址的协议。ARP 表项可以分为动态和静态两种类型。另外 ARP 还有扩展应用功能，包括 Proxy ARP 功能、免费 ARP 以及 ARPing。

## 目的

局域网中每台主机或路由器都有一个 32 位的 IP 地址，这个地址用于主机或路由器的所有通信。IP 地址的分配是独立于机器的硬件地址的。而在以太网中，主机或路由器是根据 48 位的 MAC 地址来发送、接收以太网数据帧的，这个 MAC 地址又称为物理地址或硬件地址，是制造设备时分配到以太网接口中的。因而，在实际的网络互联中，需要一种地址解析的机制来为这两种不同的地址形式提供映射。

ARP 协议主要是解决以上问题，它包括如下的应用特性：

- 动态 ARP：利用 ARP 报文，ARP 动态执行并自动进行 IP 地址到以太网 MAC 地址的解析，无需网络管理员手工处理。
- 静态 ARP：建立 IP 地址和 MAC 地址之间固定的映射关系，在主机和路由器上不能动态调整此映射关系。需要网络管理员手工添加。
- Proxy ARP 功能：
  - 路由式 Proxy ARP：当主机上没有配置缺省网关地址（即不知道如何到达本网络的中介系统），它可以发送一个 ARP 请求（请求目的主机的 MAC 地址），使能 Proxy ARP 功能的路由器收到这样的请求后，会使用自己的 MAC 地址作为该 ARP 请求的回应，使得处于不同物理网络的同一网段的主机之间可以正常的相互通信。
  - VLAN 内 Proxy ARP：如果两个用户属于相同的 VLAN，但 VLAN 内配置了用户隔离，使能了 VLAN 内 Proxy ARP 功能的接口接收到目的地址不是自己的 ARP 请求报文后，路由器并不立即丢弃该报文，而是查找该接口的 ARP 表项。如果满足代理条件，则将路由器的 MAC 地址发送给 ARP 请求方，使得相同 VLAN 内，且 VLAN 配置用户隔离后的网络上的主机之间的相互通信。
  - VLAN 间 Proxy ARP：如果两个用户属于不同的 VLAN，使能了 VLAN 间 Proxy ARP 功能的接口接收到目的地址不是自己的 ARP 请求报文后，路由器并不立即丢弃该报文，而是查找该接口的 ARP 表项。如果满足代理条件，则将路由器的 MAC 地址发送给 ARP 请求方，使得不同 VLAN 之间的主机相互通信。
- 免费 ARP：用于检查重复的 IP 地址和通告新的 MAC 地址。
- ARPing：包括 ARP-Ping IP 和 ARP-Ping MAC，这两者统称为 ARPing，用于部署二层特性时方便维护。

# 2.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC826 | Ethernet Address Resolution Protocol | |
| RFC903 | Reverse Address Resolution Protocol | |
| RFC1027 | Using ARP to Implement Transparent Subnet Gateways | |
| RFC1042 | Standard for the Transmission of IP Datagrams over IEEE 802 Networks | |

# 2.3 可获得性

## 涉及网元

无需其它网元的配合。

## License 支持

无需获得 License 许可，均可获得该特性的服务。

## 版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

## 特性依赖

不依赖其他特性。

## 硬件要求

对硬件无特殊要求。

# 2.4 原理描述

ARP 是用来实现以太网中三层 IP 地址与二层 MAC 地址之间的映射，是以太网通信的基础。

## 2.4.1 ARP 原理

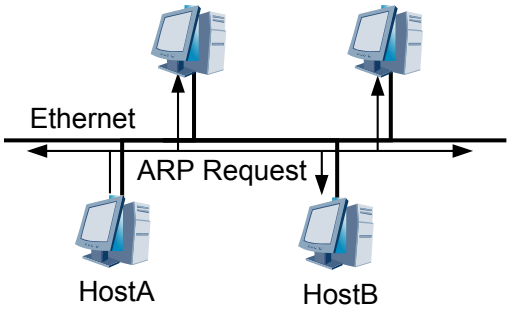
以太网的同一网段内以广播的方式查询某个 IP 地址对应的 MAC 地址，以实现三层 IP 地址与二层 MAC 地址之间的动态映射，这是任何以太网主机设备都支持的一个协议。我们有的时候称 ARP 为 2.5 层协议。

## ARP 地址解析过程

TCP/IP 协议的设计人员根据以太网这种具有广播特性的网络开发出的 ARP 地址解析协议。主机在仅知道同一物理网络上的目的端的 IP 地址情况下，通过 ARP 解析到目的端的 MAC 地址。即使网络上的主机发生变化，比如主机的增加或减少、主机更换计算机的网卡等，仍可以完成从 IP 地址到 MAC 地址的转换，并且这个转换关系可以动态更新。

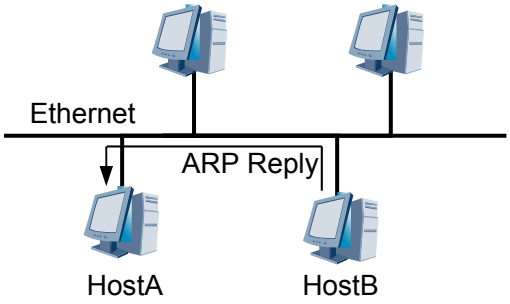
ARP 通过以下两个步骤完成地址解析过程。

**图 2-1** ARP 请求过程




如图 2-1 所示，当主机 A 仅知道主机 B 的 IP 地址而不知道其 MAC 地址时，它广播一个 ARP 请求报文，请求得到主机 B 的 MAC 地址。

**图 2-2** ARP 响应过程




因为主机 A 发送的是一个广播报文，所以包括主机 B 在内的所有主机都会接收到这个请求。由于 ARP 请求报文的目的 IP 地址是主机 B 的 IP 地址，所以只有主机 B 会响应这个 ARP 请求。如图 2-2 所示，主机 B 向主机 A 发出一个包含其 MAC 地址的 ARP 响应报文。

当主机 A 接收到主机 B 的响应报文后，就用这个 MAC 地址和主机 B 通信。

## ARP 老化机制

- 高速缓存

  如果每次主机 A 向主机 B 发送一个分组前都要发送一个广播的 ARP 请求报文的话，会增加很多网络的通信量。而且网络上的所有机器都必须接受和处理这个广播的 ARP 请求报文，这也极大的影响了网络运行效率。

为了解决以上问题，每台主机上都维护着一个高速缓存，这是 ARP 高效运行的一个关键。在这个高速缓存中，存放最近获得的 IP 地址到 MAC 地址的映射关系。

发送方在每次发送分组时，都先在缓存中查找目标 IP 地址所对应的 MAC 地址。如果 ARP 缓存中有对应的 MAC 地址，主机就不会再发送 ARP 请求报文，而是直接将分组发至这个 MAC 地址。如果 ARP 缓存中没有对应的 MAC 地址时，主机才会发送广播的 ARP 请求报文。

- 动态 ARP 表项的老化超时时间

  如图 2-2 所示，当主机 B 回应了主机 A 的 ARP 请求后，在主机 A 的缓存中会形成主机 B 的 IP 地址和其 MAC 地址的映射关系。但是，如果主机 B 发生故障后或者更换了网卡时，主机 A 没有得到任何关于主机 B 的任何通告，于是主机 A 仍会继续将分组发送给主机 B。造成地址解析出现错误的原因就是主机 A 中的缓存表的信息没有得到及时的更新。

  为了减少地址解析过程中所出现的错误，ARP 高速缓存中的表项一般都会设定一个定时器。当达到定时器的动态 ARP 表项的老化超时时间后，删除掉这个表项。

  通过设置定时器，在地址解析过程中出现错误的现象得到了改善但并没有完全消除，其原因在于时延。如果定时器的动态 ARP 表项的老化超时时间是 N 秒，发送方只有等到 N 秒后才能检测到接收方出现了故障，在此期间发送方缓存表的信息还是没有得到及时的更新。

- 动态 ARP 表项的老化探测次数

  除了设置定时器中动态 ARP 表项的老化超时时间，还可以通过设置动态的探测次数来减少地址的解析错误。在一个动态 ARP 表项老化之前，系统先进行探测，如果超过设置的探测次数后仍没有应答，则此 ARP 表项将被删除。

- 动态 ARP 表项的老化探测模式

  ARP 表项老化之前，接口会发送 ARP 老化探测报文。老化探测报文可以是单播报文，也可以是广播报文。缺省情况下，接口以广播模式发送 ARP 老化探测报文。

  当对端设备的 IP 地址不变化而 MAC 地址频繁更新时，建议使用广播模式发送 ARP 老化探测报文。

  当对端设备 MAC 地址不变，当前网络带宽资源特别紧缺，且 ARP 表项的老化时间设置的比较小时，建议使用单播模式发送 ARP 老化探测报文。

  当其他厂商设备与华为设备互联时，其他厂商设备接收到目的 MAC 地址为广播地址的 ARP 老化探测报文后，若 ARP 表项中已存在华为设备的 IP 地址与 MAC 地址映射，则不响应该广播 ARP 老化探测报文。这种特殊情况下华为设备需要配置成以单播方式发送 ARP 老化探测报文，即单播的 ARP 老化探测模式。

## 2.4.2 动态 ARP

### ARP 表项的创建与更新

依据 ARP 协议描述，几乎所有的以太网通信都以 ARP 开始，所以任何以太网主机设备都支持这个协议，而且 IP 地址到以太网 MAC 地址的解析主要也是动态生成，无须网络管理员手工处理。

一般实现中，如果收到的 ARP 报文满足以下条件中的任何一条，系统将创建或更新 ARP 表项：

- ARP 报文的源 IP 地址与入接口 IP 地址在同一网段，且不是广播地址，目的 IP 地址是本接口 IP 地址。

- ARP 报文的源 IP 地址与入接口 IP 地址在同一网段，且不是广播地址，目的 IP 地址是本接口的 VRRP（Virtual Router Redundancy Protocol）虚拟 IP 地址。
- ARP 报文的源 IP 地址与入接口 IP 地址在同一网段，且不是广播地址，入接口是 IPoEoA 应用的 Virtual-Ethernet 接口。
- ARP 报文的目的 IP 地址是入接口上配置的 NAT 地址池中的地址。

如果收到的 ARP 报文的源 IP 地址在入接口的 ARP 表中已经存在对应表项，也将对 ARP 表项进行更新。

### ARP 抑制功能

在特殊组网或者遭受到 ARP 攻击时，系统在同一时间内会接收到多个源 IP 地址相同的 ARP 报文，这就需要系统对 ARP 表项进行重复更新。为了维护系统性能，系统可以启动 ARP 抑制功能，即在 1s 内收到多次源 IP 地址相同的 ARP 报文，系统将只通知发送 ARP 报文的设备已收到 ARP 报文，而不更新设备的 ARP 表。

如果对所有接口都做 ARP 抑制会造成某些接口的 ARP 表项暂时无法正常更新。ARP 抑制只针对 VLANIF 和 Eth-Trunk 两类接口，缺省情况下，对 VLANIF 始终会进行抑制，其它逻辑接口进行可配抑制。

### 二层拓扑探测功能

二层拓扑探测功能，是指当二层接口的状态由 Down 变为 Up 时，二层接口所属 VLAN 对应的所有 ARP 表项的老化超时时间变为 0，使设备重新发送 ARP 探测报文，更新二层接口所属 VLAN 对应的所有 ARP 表项。

## 2.4.3 静态 ARP

静态 ARP 是指 IP 地址和 MAC 地址之间有固定的映射关系，在设备上不能动态调整此映射关系。

配置静态 ARP 表项增加通信的安全性。配置静态 ARP 表项可以限制与指定 IP 地址的设备通信时只使用指定的 MAC 地址，此时攻击报文无法修改此表项的 IP 地址和 MAC 地址的映射关系，从而保护了本设备和指定设备间的正常通信。

静态 ARP 表项通过手工配置和维护，不会被老化，不会被动态 ARP 表项覆盖。

## 2.4.4 Proxy ARP

Proxy ARP 主要是通过代理的方式来解决网络互通问题的 ARP 实现功能。

Proxy ARP 有以下特点：

- 所有处理在 ARP 子网网关（ARP Subnet Gateways）进行，所连网络中的主机不必做任何改动；
- 在主机端看不到子网，只是一个标准 IP 网络；
- Proxy ARP 只影响主机的 ARP 高速缓存，对网关的 ARP 高速缓存和路由表没有影响；
- 使用 Proxy ARP 后，主机应该减小 ARP 老化时间，以尽快使无效 ARP 项失效，减少发给路由器而路由器却不能转发的报文。

下表为三种 Proxy ARP：

| Proxy ARP 方式 | 解决的问题 |
| --- | --- |
| 路由式 Proxy ARP | 解决同一网段不同物理网络上计算机的互通问题。 |
| VLAN 内 Proxy ARP | 解决相同 VLAN 内，且 VLAN 配置用户隔离后的网络上计算机互通问题。 |
| VLAN 间 Proxy ARP | 解决不同 VLAN 之间对应计算机的三层互通问题。 |

### 路由式 Proxy ARP

路由式 Proxy ARP 就是使那些在同一网段却不在同一物理网络上的计算机或路由器能够相互通信的一种功能。

在实际应用中，如果连接路由器的当前主机上没有配置缺省网关地址（即不知道如何到达本网络的中介系统），此时将无法进行数据转发。

路由式 Proxy ARP 可以解决这个问题，主机发送一个 ARP 请求（请求目的主机的 MAC 地址），使能 Proxy ARP 功能的路由器收到这样的请求后，会使用自己的 MAC 地址作为该 ARP 请求的回应，以此进行数据转发。

使能 Proxy ARP 功能的路由器还可隐藏物理网络的细节，使得处于不同物理网络但网络号相同的两个 Ethernet 的内部主机之间可以正常的相互通信。

### VLAN 内 Proxy ARP

如果两个用户属于相同的 VLAN，但 VLAN 内配置了用户隔离。此时用户间要互通，需要在关联了 VLAN 的接口上启动 VLAN 内 Proxy ARP 功能。

若路由器的接口使能了 VLAN 内 Proxy ARP 功能，接口在接收到目的地址不是自己的 ARP 请求报文后，路由器并不立即丢弃该报文，而是查找该接口的 ARP 表项。如果满足代理条件，则将路由器的 MAC 地址发送给 ARP 请求方。

VLAN 内 Proxy ARP 主要用于配置了用户隔离的 VLAN 内的用户间互通。

### VLAN 间 Proxy ARP

如果两个用户属于不同的 VLAN，用户间要进行三层互通，可以在关联了 VLAN 的接口上启动 VLAN 间 Proxy ARP 功能。

若路由器的接口使能了 VLAN 间 Proxy ARP 功能，接口在接收到目的地址不是自己的 ARP 请求报文后，路由器并不立即丢弃该报文，而是查找该接口的 ARP 表项。如果满足代理条件，则将路由器的 MAC 地址发送给 ARP 请求方。

VLAN 间 Proxy ARP 主要用于：

- 处于不同 VLAN 的用户进行三层通信。
- 可在 Super VLAN 对应的 VLANIF 接口上启动 VLAN 间 Proxy ARP 功能，实现 Sub VLAN 间用户互通。

## 2.4.5 免费 ARP

主机主动使用自己的 IP 地址作为目标地址发送 ARP 请求，此种方式称免费 ARP。免费 ARP 有三方面的作用：

- 用于检查重复的 IP 地址：正常情况下不会收到 ARP 回应，如果收到，则表明本网络中存在与自身 IP 地址重复的地址。
- 用于通告一个新的 MAC 地址：发送方更换了网卡，MAC 地址变了，为了能够在 ARP 表项老化前通告所有主机，发送方可以发送一个免费 ARP。
- 在 VRRP 备份组中用来通告主备发生变换。

## 2.4.6 ARP-Ping

ARP-Ping：包括 ARP-Ping IP 和 ARP-Ping MAC，用于部署二层特性时方便维护。

### ARP-Ping IP

ARP-Ping IP 的标准是 ARP 协议。通过配置管理平面获得用户输入的 IP 地址和出接口（出接口是可选项），然后构造 ARP Request 报文，在出接口广播该报文。在指定超时时间内，若收不到回复报文，则向用户显示该 IP 地址无人使用；若收到回复报文，则将回复报文中的对端 MAC 地址取出，显示给用户。

ARP-Ping IP 是利用 ARP 报文在局域网内探测 IP 地址是否被其它的设备使用的一种方法。

用户对设备配置 IP 地址前，需要确认该 IP 地址有没有被网络上的其他设备使用，可以通过发送 ARP 报文（二层），确认该 IP 的使用情况，以便做出相应调整。

通过 **ping** 命令也可以探测该 IP 地址是否被网络上的其他设备使用。但是如果带有防火墙功能的目的主机和路由设备设置了对 Ping 报文不进行回复的功能时，就不会响应 Ping 报文，造成该 IP 没有被使用的假象。由于 ARP 报文是二层协议，大多数情况下可以透过设置了对 **ping** 报文不进行回复的防火墙，从而避免了此类情况的发生。

**ARP-Ping IP 原理**

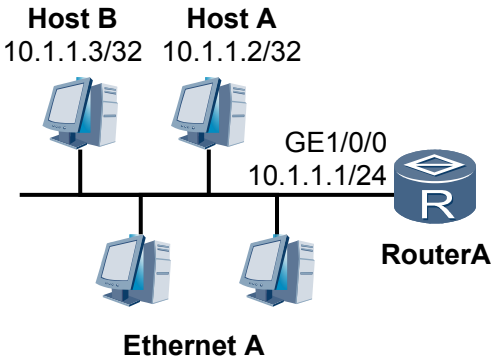
ARP-Ping IP 发送的是 ARP 请求报文。以下是 ARP-Ping IP 的具体实现过程。

1. 用户通过命令行设置指定的 IP 地址后，发送 ARP 请求报文并且启动 ARP Reply 报文的超时定时器。
2. 局域网内路由设备或主机收到 ARP 请求报文后，回复 ARP Reply 报文。
3. 源路由设备收到 ARP Reply 报文后将 Reply 报文中的源 IP 地址和命令行中输入的 IP 地址进行比较。若匹配，则向用户显示与所输入 IP 地址相对应的 MAC 地址并且关闭 Reply 报文的超时定时器，本次操作结束。

   若 ARP Reply 报文的超时定时器超时，输出该 IP 地址无设备使用的显示信息。

如图 2-3 所示，RouterA 可通过 ARP-Ping IP 来获知 10.1.1.2 这个 IP 地址是否被使用。RouterA 收到网络内 IP 地址为 10.1.1.2 的主机 A 的 ARP Reply 报文后，将这个主机的 MAC 地址显示出来。通过显示信息可得知这个 IP 地址被网络内的主机使用。

**图 2-3** ARP-Ping IP 的实现过程




## ARP-Ping MAC

ARP-Ping MAC 和普通 Ping 处理一样，但 ARP-Ping MAC 只应用在直连以太局域网或二层 VPN 以太网络。发送 ICMP 回显请求报文，接收 ICMP Reply 报文，解析报文，把保存在报文数据区的源 MAC 地址和本机保存的 MAC 地址相比较。如果相同则显示该报文的 IP 地址，并提示该 MAC 地址已被使用，关闭超时定时器，本次操作结束。当 ICMP Request 报文响应时间超时，输出该 MAC 地址无设备使用的信息。

### ARP-Ping MAC 的基本原理

ARP-Ping MAC 发送的是广播的 ICMP 请求（ECHO Request）报文。以下是 ARP-Ping MAC 的具体实现过程：

1. 用户从命令行通过设置指定的 MAC 地址后，发送广播的 ICMP 报文并且启动超时定时器。
2. 局域网内各路由设备或主机收到 ICMP 请求报文后，回复 ICMP 应答（ECHO Reply）报文。
3. 源路由设备收到 ICMP 应答报文后，将 Reply 报文中的源 MAC 地址和命令行中输入的 MAC 地址相比较。若匹配，显示出该报文的 IP 地址，并提示该 MAC 地址已被使用并且关闭超时定时器，本次操作结束。

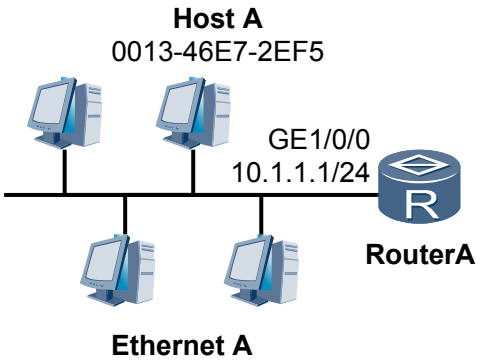
   若 ICMP 应答报文的响应时间定时器超时，输出该 MAC 地址无设备使用的信息。

说明

如果系统关闭了回复网段地址的报文请求，发送方是收不到 ICMP 响应报文的。

如图 2-4 所示，RouterA 可通过 ARP-Ping MAC 来获知 0013-46E7-2EF5 这个 MAC 地址是否被使用。RouterA 收到网络内所有主机回复的 ICMP 的响应报文后，将 MAC 地址为 0013-46E7-2EF5 的主机的 IP 地址显示出来。通过显示信息可得知这个 MAC 地址所对应的 IP 地址。

**图 2-4** ARP-Ping MAC 的实现过程

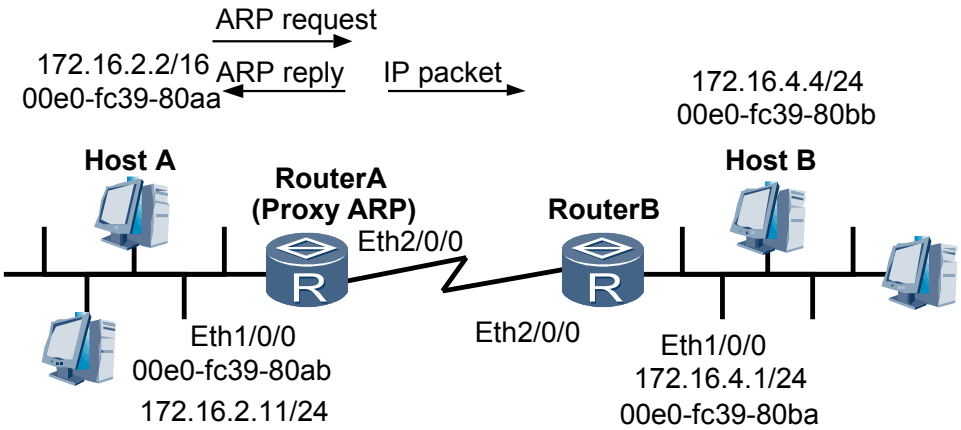



# 2.5 应用

## 路由式 Proxy ARP

如图 2-5 所示，由于 Host A 只有 16 位掩码，它认为自己直连到 172.16.0.0 网段，当 Host A 需要与 172.16.0.0 网段上的设备例如 Host B 通信时，它发送 ARP 报文请求 Host B 的物理地址。由于 Host B 和 Host A 不在同一个广播域中，所以 Host B 收不到 Host A 发出的 ARP 请求。

**图 2-5** 路由式 Proxy ARP 典型组网图

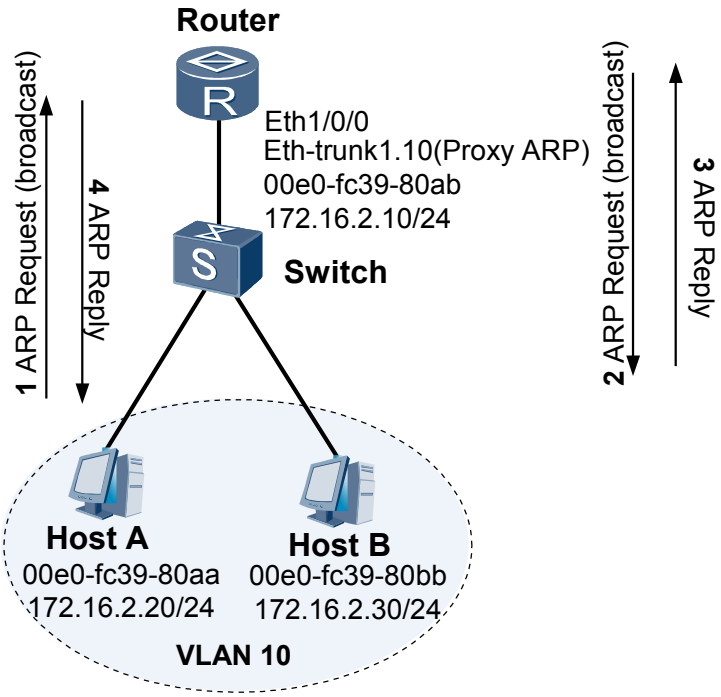



如果在 RouterA 的接口 Eth1/0/0 上使能路由式 Proxy ARP，由 RouterA 转发 Host A 与 Host B 之间的 IP 报文，Host A 就可以与 Host B 互通了。

## VLAN 内 Proxy ARP

如图 2-6 所示，Host A 和 Host B 是交换机设备下的两个用户。连接 Host A 和 Host B 的两个接口在交换机设备上属于同一个 VLAN10。由于在交换机设备上配置了 VLAN 内不同接口彼此隔离，因此 Host A 和 Host B 不能直接在二层互通。

**图 2-6** VLAN 内 Proxy ARP 典型组网图

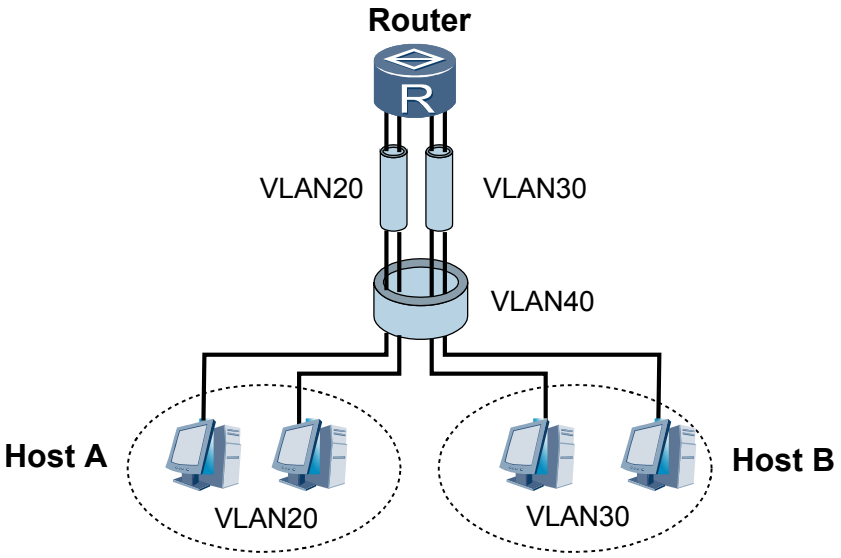



如果在 Router 上创建子接口 Eth-trunk1.10，使子接口关联 VLAN10。在 Router 的子接口 Eth-trunk1.10 上使能 VLAN 内 Proxy ARP，Host A 和 Host B 就可以在二层互通了。子接口 Eth-trunk1.10 的 IP 地址与 VLAN10 中的主机 IP 地址必须在同一个网段。

## VLAN 间 Proxy ARP

如图 2-7 所示，Host A 和 Host B 是 Router 下的两个用户。由于连接 Host A 和 Host B 的两个接口在 Router 上属于不同的 VLAN，因此 Host A 和 Host B 不能直接实现二层互通。

**图 2-7** VLAN 间 Proxy ARP 典型组网图



如果在 AR200-S 上创建 SuperVLAN 40，将 VLAN20 和 VLAN30 加入 VLAN40，并且创建接口 VLANIF 40，在 VLANIF 40 上使能 VLAN 间 Proxy ARP，Host A 和 Host B 就可以实现三层互通了。VLANIF 40 的 IP 地址与 VLAN20 和 VLAN30 中的主机 IP 地址在同一个网段。

# 2.6 术语与缩略语

## 缩略语

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| ARP | Address Resolution Protocol | 地址解析协议 |
| VRRP | Virtual Router Redundancy Protocol | 虚拟路由冗余协议 |
| VLAN | Virtual Local Area Netw | 虚拟局域网 |

# 3 IPv4

## 关于本章

# 3.1 介绍

## 定义

IPv4（Internet Protocol Version 4）是 TCP/IP 协议族中最为核心的协议。它工作在 TCP/IP 协议栈的互联网络层。该层与 OSI 参考模型的网络层相对应。IP 层提供了无连接数据传输服务，即将信息分割成数据单元，以数据报的形式从网络的一个地方传送到另一个地方。

## 目的

屏蔽各链路层差异，为上层提供统一的网络层传输标准。

# 3.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC793 | Transmission Control Protocol | |
| RFC768 | User Datagram Protocol | |

# 3.3 可获得性

## 涉及网元

无需其他网元的配合。

## License 支持

无需获得 License 许可，即可获得该特性的服务。

## 版本支持

**表 3-1** 最低版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

## 特性依赖

不依赖其他特性。

### 硬件要求

对硬件无特殊要求。

## 3.4 原理描述

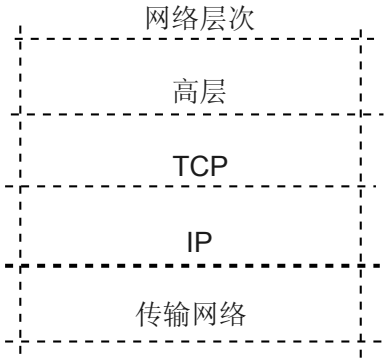
### 3.4.1 TCP 原理描述

传输控制协议（TCP）由 RFC793 定义，用于在主机间实现面向连接的可靠性服务。TCP 协议为用户进程定义了一个可靠的、面向连接的、全双工的服务。

TCP 是面向连接的端到端的可靠协议。它支持多种网络应用程序。TCP 假定下层只能提供不可靠的数据报服务，它可以在多种硬件构成的网络上运行。

图 3-1 表示了 TCP 在层次式结构中的位置，它的下层是 IP 协议，TCP 可以根据 IP 协议提供的服务传送大小不定的数据，IP 协议负责对数据进行分段、重组，在多种网络中传送。

**图 3-1 层次式结构**



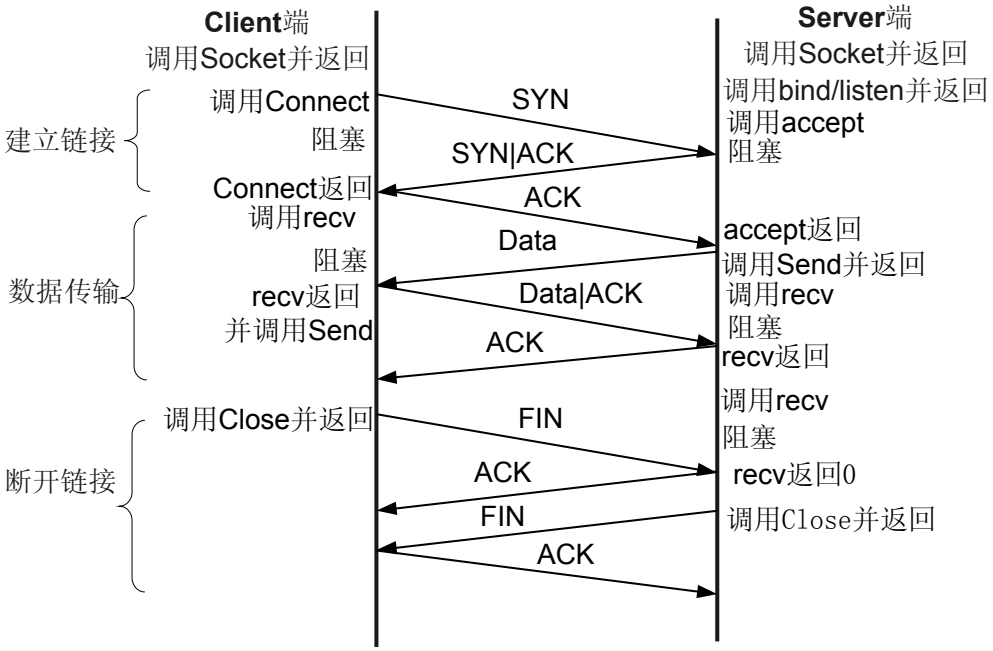

在 ISO 层次结构中，TCP 的上层是应用程序，下层是 IP 协议。

对于上层应用程序，TCP 应该能够异步传送数据。下层接口假定为 IP 协议接口。为了在并不可靠的网络上实现面向连接的可靠的传送数据，TCP 必须：

- 解决可靠性、流量控制的问题
- 为上层应用程序提供多个接口
- 为多个应用程序提供数据
- 解决连接问题
- 解决通信安全性的问题

图 3-2 表示了 TCP 连接建立和拆除过程。

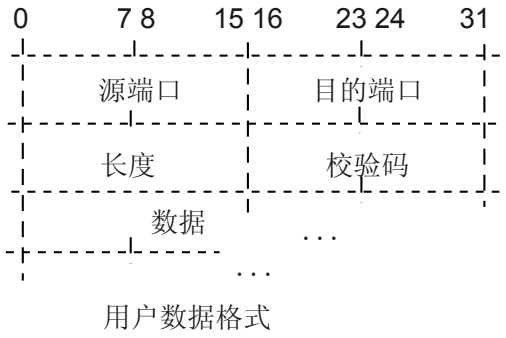
**图 3-2** TCP 连接建立和拆除过程




## 3.4.2 UDP 原理描述

UDP（用户数据报协议）是用来在互连网络环境中提供包交换的计算机通信协议。此协议默认为网际协议（IP）是其下层协议，提供了向另一用户程序发送信息的最简便的协议机制。UDP 是面向操作的，未提供数据提交和复制保护。如果应用程序要求可靠的数据传送应该使用传输控制协议（TCP）。数据报格式如图 3-3 所示。

**图 3-3** UDP 协议报文格式



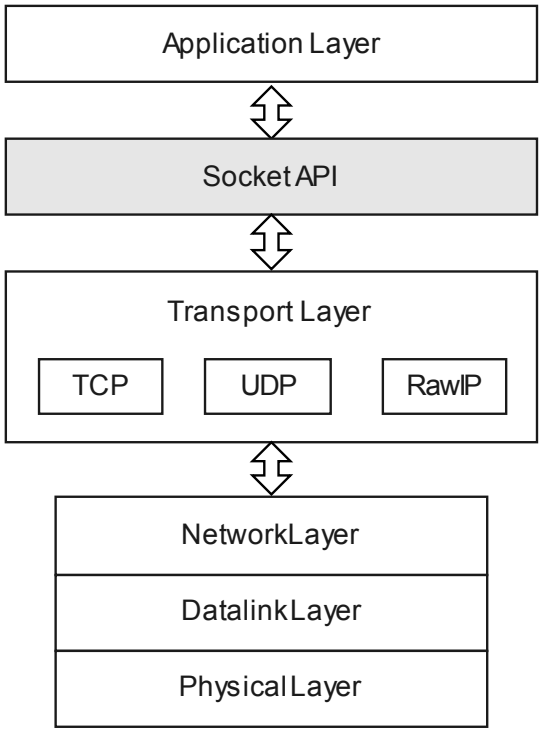

## 3.4.3 RawIP 原理描述

RawIP 只填充 IP 首部的有限几个字段，但它允许应用进程提供自己的 IP 首部。它与 UDP 类似，是不可靠的，即没有任何控制能确定 RawIP 数据报是否已被接收。它是无连接的，即在主机间传输数据时，不需要任何类型的电路。RawIP 相比 UDP 的区别在于：RawIP 允许应用程序直接通过 Socket 接口操作 IP 层。对于许多需要跟下层直接交互的应用，非常方便。

### 3.4.4 Socket 原理描述

Socket 是一组编程接口（API），介于传输层与应用层之间，屏蔽传输层差异，向应用层提供统一的编程接口。应用层可以不必了解 TCP/IP 协议的细节，直接通过对 Socket 接口函数的调用完成数据在 IP 网络中的传输。图 3-4 表示了 Socket 在 TCP/IP 协议栈中的位置。

**图 3-4** Socket 分层模型




基于传输层差异，目前支持四种类型的 Socket：

- 基于 TCP 的 Socket，向应用层提供的一种可靠的流式数据通讯服务。
- 基于 UDP 的 Socket，向应用层提供给一种无连接的，不可靠的数据传输，但是这种基于数据报的传输可以提供报文边界。
- 基于 RawIP 的 Socket，也叫 Raw Socket。与基于 UDP 的 Socket 类似，也是无连接的，不可靠的数据传输，同样可以提供报文边界。但是它的特点是能够使应用程序直接访问网络层。
- 基于链路层的 Socket，这是为 IS-IS 路由协议提供的 Socket 接口，使 IS-IS 路由协议可以通过该 Socket 接口直接访问链路层。

## 3.5 应用

### ICMP 报文发送开关

在正常情况下，设备可以正确发送 ICMP 主机不可达报文和 ICMP 重定向报文。但是，当网络流量较大时，则设备会发送大量的 ICMP 报文，增大网络的流量负担。同时，

ICMP 差错报文经常被利用发起网络攻击，容易产生恶性循环，从而加剧网络的拥塞。如重定向报文可以被用来使路由频繁变更。

AR200-S 提供在 ICMP 报文的出接口增加两个控制开关，分别用来打开或关闭 ICMP 主机不可达报文和 ICMP 重定向报文的发送开关。如果关闭这两个开关，则路由设备不会发送这两种报文，从而起到减小网络流量、降低设备负担、防止遭到恶意攻击的作用。

# 3.6 术语与缩略语

## 术语

无

## 缩略语

| 缩略语 | 全称 |
| --- | --- |
| TCP | Transmission Control Protocol |
| UDP | User Datagram Protocol |

# 4 IPv6

## 关于本章

本章介绍 IPv6 原理和应用。

# 4.1 介绍

## 定义

IPv6（Internet Protocol Version 6）是网络层协议的第二代标准协议，也被称为 IPng（IP Next Generation）。它是 IETF（Internet Engineering Task Force，Internet 工程任务组）设计的一套规范，是 IPv4（Internet Protocol Version 4）的升级版本。IPv6 和 IPv4 之间最显著的区别就是 IP 地址长度从原来的 32 位升级为 128 位。IPv6 以其简化的报文头格式、充足的地址空间、层次化的地址结构、灵活的扩展头、增强的邻居发现机制将在未来的市场竞争中充满活力。

AR200-S 支持在下列接口配置 IPv6 功能：

- Ethernet 接口及子接口
- Serial 接口（只有 link-protocol 为 PPP 或 HDLC 的 Serial 接口支持 IPv6 功能）
- Tunnel 接口
- Loopback 接口
- Eth-Trunk 接口、Eth-Trunk 子接口
- VLANIF 接口

## 目的

以 IPv4 为核心技术的 Internet 获得巨大成功，促使 IP 技术得到广泛应用。然而，随着因特网的迅猛发展，IPv4 设计的不足也日益明显，主要有以下几点：

- IPv4 地址空间不足

  IPv4 地址采用 32 比特标识，理论上能够提供的地址数量是 43 亿。但由于地址分配的原因，实际可使用的数量不到 43 亿。另外，IPv4 地址的分配也很不均衡：美国占全球地址空间的一半左右，而欧洲则相对匮乏；亚太地区则更加匮乏。与此同时，移动 IP 和宽带技术的发展需要更多的 IP 地址。IPv4 地址资源紧张直接限制了 IP 技术应用的进一步发展。

  针对 IPv4 的地址短缺问题，也曾先后出现过几种解决方案。比较有代表性的是 CIDR(Classless Inter-Domain Routing)和 NAT(IP Network Address Translator)。但是 CIDR 和 NAT 都有各自的弊端和不能解决的问题，由此推动了 IPv6 的发展。

- 骨干设备维护的路由表表项数量过大

  由于 IPv4 发展初期的分配规划问题，造成许多 IPv4 地址分配不连续，不能有效聚合路由。日益庞大的路由表耗用较多内存，对设备成本和转发效率产生影响，这一问题促使设备制造商不断升级其产品，以提高路由寻址和转发性能。

- 不易进行自动配置和重新编址

  由于 IPv4 地址只有 32 比特，并且地址分配不均衡，导致在网络扩容或重新部署时，经常需要重新分配 IP 地址。因此需要能够进行自动配置和重新编址以减少维护工作量。

- 不能解决日益突出的安全问题

  随着因特网的发展，安全问题越来越突出。IPv4 协议制定时并没有仔细针对安全性进行设计，因此固有的框架结构并不能支持端到端的安全。IPv6 将 IPSec 作为它的标准扩展头实现，可以提供端到端的安全特性。

IPv6 技术从根本上解决了 IP 地址短缺的问题；且易于部署，能够兼容当前的各种应用，方便用户的平滑过渡；同时可实现与 IPv4 网络的共存和互通。由于 IPv4 存在以上种种弊端和不足，IPv6 技术的优越性显而易见，因此 IPv6 技术得以迅猛发展。

# 4.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC793 | Transmission Control Protocol | |
| RFC768 | User Datagram Protocol | |
| RFC1981 | Path MTU Discovery for IP version 6 | |
| RFC2460 | Version 6 of the Internet Protocol (IPv6), also sometimes referred to as IP Next Generation or IPng. | |
| RFC2461 | Neighbor Discovery for IP Version 6 (IPv6) | |
| RFC2463 | Internet Control Message Protocol for the Internet Protocol Version 6 Specification | |
| RFC2465 | Management Information Base for IP Version 6:<br>Textual Conventions and General Group | |
| RFC2466 | Management Information Base for IP Version 6:<br>ICMPv6 Group | |
| RFC2473 | Generic Packet Tunneling in IPv6 Specification | |
| RFC2711 | IPv6 Router Alert Option | |
| RFC2893 | Transition Mechanisms for IPv6 Hosts and Routers | |
| RFC3056 | Connection of IPv6 Domains via IPv4 Clouds | |
| RFC3068 | An Anycast Prefix for 6to4 Relay Routers | |
| RFC3484 | Default Address Selection for Internet Protocol Version 6 (IPv6) Section 2.1 | |

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC3971 | SEcure Neighbor Discovery (SEND) | |
| RFC3972 | Cryptographically Generated Addresses (CGA) | |
| RFC4191 | Default Router Preferences and More-Specific Routes | |
| RFC4214 | Intra-Site Automatic Tunnel Addressing Protocol(ISATAP) | |
| RFC4291 | Internet Protocol Version 6 (IPv6) Addressing Architecture | |
| RFC4443 | Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification | |
| RFC4861 | Neighbor Discovery for IP version 6 (IPv6) | |

# 4.3 可获得性

## 涉及网元

无需其他网元的配合。

## License 支持

无需获得 License 许可，即可获得该特性的服务。

## 版本支持

**表 4-1** 最低版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

## 特性依赖

不依赖其他特性。

## 硬件要求

对硬件无特殊要求。

# 4.4 原理描述

IPv6 基本功能主要包括 IPv6 邻居发现、IPv6 路径 MTU 发现。邻居发现和 Path MTU 发现机制均是基于 ICMPv6 协议报文实现的。

## 4.4.1 IPv6 地址

### IPv6 地址的书写格式

IPv6 的 128 位 IP 地址有以下两种表示形式。

- X:X:X:X:X:X:X:X
  - 在这种形式中，128 位的 IPv6 地址被分为 8 组，每组的 16 位用 4 个十六进制字符（0 ～ 9，A ～ F）来表示，组和组之间用冒号（:）隔开。其中每个“X”代表一组十六进制数值。比如下面这个 IPv6 地址：

    2031:0000:130F:0000:0000:09C0:876A:130B

    为了书写方便，每组中的前导“0”都可以省略，所以上述地址可写为：

    2031:0:130F:0:0:9C0:876A:130B。
  - 另外，地址中包含的连续两个或多个均为 0 的组，可以用双冒号“::”来代替，这样可以压缩 IPv6 地址书写时的长度，所以上述地址又可以进一步简写为：

    2031:0:130F::9C0:876A:130B。

    在一个 IPv6 地址中只能使用一次双冒号“::”，否则当计算机将压缩后的地址恢复成 128 位时，无法确定每段中 0 的个数。
- X:X:X:X:X:X:d.d.d.d
  - 分为如下两种类型：
    - IPv4 兼容 IPv6 地址。地址格式为：0:0:0:0:0:0:IPv4-address，其高阶 96bits 均为 0，其低阶 32bits 是一个 IPv4 地址。该 IPv4 地址必须是 IPv4 网络中可达的 IPv4 地址，且不能是组播地址、广播地址、环回地址或未指定的地址（0.0.0.0）。
    - IPv4 映射 IPv6 地址。地址格式为：0:0:0:0:0:FFFF:IPv4-address。该地址用来将 IPv4 节点的地址表示为 IPv6 地址。

  其中 IPv4 兼容 IPv6 地址用于配置 IPv6 over IPv4 隧道。

  其中“X:X:X:X:X:X”代表高阶的六组数字，用十六进制数来表示每组的 16 比特。“d”代表低阶的四组数字，用十进制数表示每组的 8 比特。后边的部分（d.d.d.d）其实就是一个标准的 IPv4 地址。

  说明

  AR150/200 暂不支持 IPv6 over IPv4 隧道功能和 IPv4 over IPv6 隧道功能。

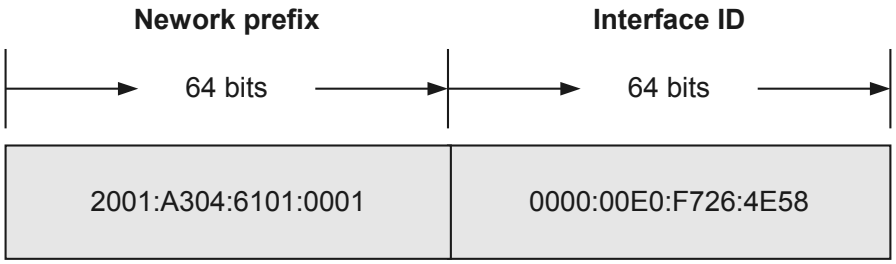
### IPv6 地址的结构

一个 IPv6 地址可以分为如下两部分：

- 网络前缀：n 比特，相当于 IPv4 地址中的网络 ID
- 接口标识：128-n 比特，相当于 IPv4 地址中的主机 ID

地址 2001:A304:6101:1:0000:E0:F726:4E58 /64 的构成如图 4-1 所示。

**图 4-1** 地址 2001:A304:6101:1:0000:E0:F726:4E58 /64 的构成示意图




## IPv6 的地址分类

IPv6 主要有三种地址：

- 单播地址（Unicast）：唯一标识一个接口，类似于 IPv4 的单播地址。发送到单播地址的数据包将被传输到此地址所标识的唯一接口。

  单播地址还可以分为四种，如表 4-2 所示。

  **表 4-2** IPv6 单播地址类型

| 地址类型 | 二进制前缀 | IPv6 前缀标识 |
| --- | --- | --- |
| 链路本地单播地址 | 1111111010 | FE80::/10 |
| 环回地址 | 00...1 (128 bits) | ::1/128 |
| 未指定地址 | 00...0 (128 bits) | ::/128 |
| 全球单播地址 | 其他 | - |

  表中各类地址的意义如下：

  - 链路本地单播地址：用于邻居发现协议和无状态自动配置进程中链路本地上节点之间的通信。使用链路本地地址作为源或目的地址的数据包不会被转发到其他链路上。使用链路本地前缀 FE80::/10(1111 1110 10)和 IEEE EUI-64 格式的接口标识符（EUI-64 可来源于 EUI-48）可在任意接口对其进行自动配置。
  - 环回地址 0:0:0:0:0:0:0:1 或::1，不会被分配给任何接口。它的作用与在 IPv4 中的 127.0.0.1 相同，即节点将 IPv6 报文发送给自己。
  - 未指定地址（::），不能被分配给任何节点，也不能作为目的地址。在主机初始化且没有取得自己的地址时，未指定地址可以用在 IPv6 报文的源地址字段，例如重复地址探测时，NS 报文的源地址就是未指定地址。
  - 全球单播地址等同于 IPv4 公网地址。用于可以聚合的链路，最后提供给网络服务提供商。这种地址类型的结构允许路由前缀的聚合，从而满足全球路由表项的数量限制。地址包括运营商管理的 48 位路由前缀和本地站点管理的 16 位子网 ID，以及 64 位接口 ID。如无特殊说明，全球单播地址包括站点本地单播地址。

- 任播地址（Anycast）：用来标识一组接口（通常这组接口属于不同的节点）。发送到任播地址的数据包被传输给此地址所标识的一组接口中距离源节点最近的一个接口（最“近”的一个，是指根据路由协议的距离度量）。

应用场合：当移动主机需要与它的“home”子网上的移动代理之一通信时，它将用该子网路由设备的任播地址。

具体地址规定：任播地址没有独立的地址空间，它们可使用任何单播地址的格式。因此，需要一种语法来区别任播地址和单播地址。

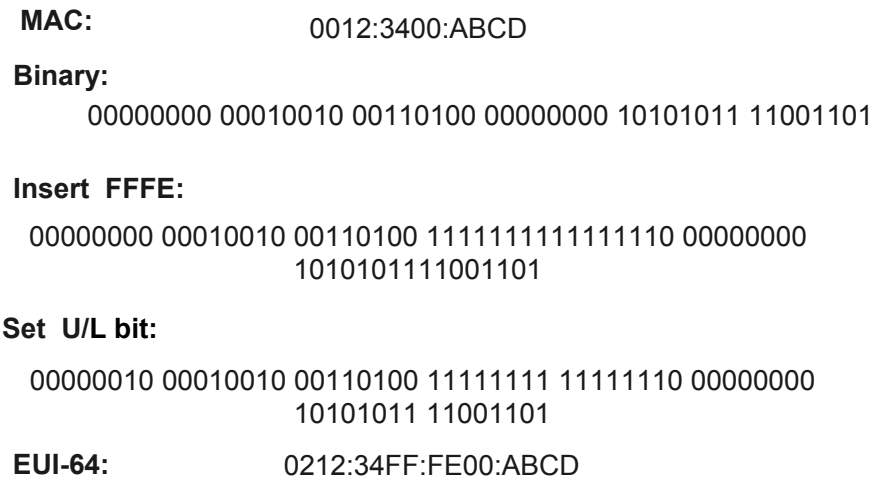
- 组播地址（Multicast）：用来标识属于不同节点的一组接口，类似 IPv4 的组播地址。发送到组播地址的数据包被传输给此地址所标识的所有接口。

  IPv6 不包括广播地址，广播地址的功能均由组播地址来提供。

### IEEE EUI-64 格式的接口标识符

IPv6 地址中的 64 位接口标识符（Interface ID）用来标识链路上的唯一接口。这个地址是从接口的链路层地址（如 MAC 地址）变化而来的。IPv6 地址中的接口标识符是 64 位，而 MAC 地址是 48 位，因此需要在 MAC 地址的中间位置插入十六进制数 FFFE（1111 1111 1111 1110）。然后将 U/L 位（从高位开始的第 7 位）设置为“1”，这样就得到了 EUI-64 格式的接口 ID。具体转换过程如图 4-2。

**图 4-2** MAC 地址到 EUI-64 格式的转换过程




## 4.4.2 IPv6 的特点

- 层次化的地址结构

  IPv6 的地址空间采用了层次化的地址结构，利于路由快速查找，同时借助路由聚合，可减少 IPv6 路由表的大小，提高路由设备的转发效率。

- 地址自动配置

  为了简化主机配置，IPv6 支持有状态地址配置（Stateful Address Autoconfiguration）和无状态地址配置（Stateless Address Autoconfiguration）。

  - 对于有状态地址配置，主机通过服务器获取地址信息和配置信息。
  - 对于无状态地址配置，主机自动配置地址信息，地址中带有本地路由设备通告的前缀和主机的接口标识。如果链路上没有路由设备，主机只能自动配置链路本地地址，实现与本地节点的互通。

- 源/目的地址选择

当网络管理者需要指定和预知系统发送报文的源/目的地址时，可以定义一组地址选择规则，这些规则构成地址选择策略表。该表类似于路由表，使用最长匹配原则查找规则。地址选择的结果是由源地址和目的地址共同决定的。

依次根据以下规则进行源地址选择，规则的编号越小，优先级越高。

1. 源地址和目的地址相同
2. 合适的生效范围
3. 避免使用已经废弃的地址
4. 家乡地址（home address）
5. 出接口地址
6. 源地址的 *label* 值和目的地址的 *label* 值相同
7. 最长匹配原则

依次根据以下规则进行目的地址选择，规则的编号越小，优先级越高。

1. 避免使用不可用的目的地址
2. 合适的生效范围
3. 避免使用已经废弃的地址
4. 家乡地址（home address）
5. 目的地址的 *label* 值和源地址的 *label* 值相同
6. 较高的 *precedence* 值
7. 在本地转发报文，不需要使用 6over4 或 6to4 隧道
8. 更小的生效范围
9. 最长匹配原则
10. 遵循原来的顺序

说明

AR150/200 暂不支持 IPv6 策略路由功能。

- 支持 QoS

  IPv6 报头的新字段定义了流量应该被如何标识和处理。通过报文头里的流标签（Flow Label）字段完成流量标识，允许路由设备对某一流中的报文进行识别并提供特殊处理。

  由于 IPv6 报头可对流量进行识别，即使是带有 IPSec 加密的报文载荷也可对其 QoS 进行保证。

  说明

  AR150/200 暂不支持基于 IPv6 的 QoS 特性。

- 内置安全性

  IPv6 将 IPSec 作为它的扩展报头实现，提供端到端的安全特性。这一特性为解决网络安全问题提供了标准，并提高了不同 IPv6 实现的互操作性。

  说明

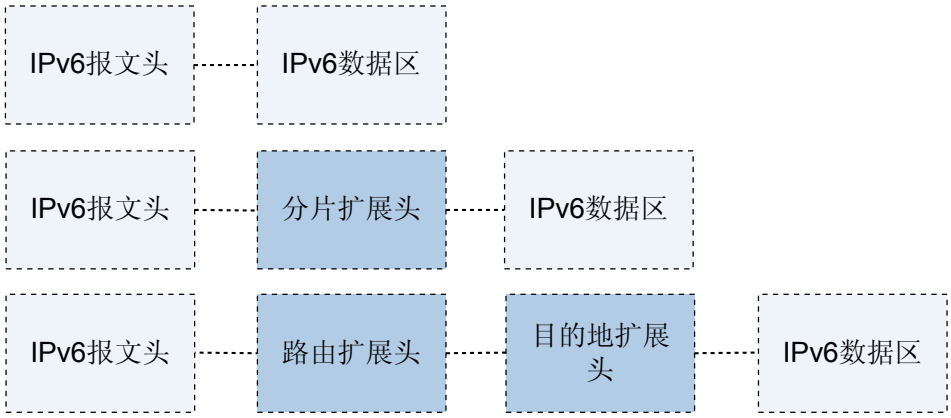
  AR150/200 暂不支持 IPv6 的内置安全特性。

- 灵活的扩展报文头

  IPv4 报头只能支持 40 字节的选项，而 IPv6 扩展报头的大小只受到 IPv6 报文大小的限制。

  IPv6 取消了 IPv4 报头中的选项字段，并引入了多种扩展报文头，在提高处理效率的同时还增强了 IPv6 的灵活性，为 IP 协议提供了良好的扩展能力。如图 4-3 所示。

**图 4-3** IPv6 扩展报文头




当超过一种扩展报头被用在同一个分组里时，报头必须按照下列顺序出现：

- IPv6 基本报头
- 逐跳选项扩展报头
- 目的选项扩展报头
- 路由扩展报头
- 分片扩展报头

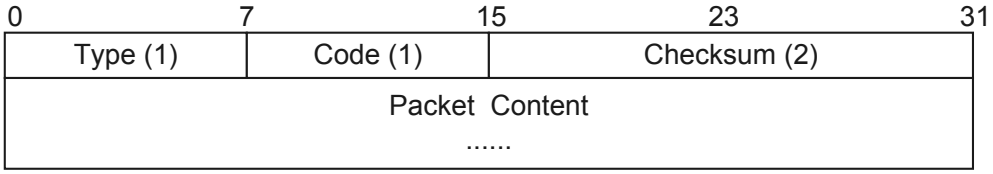
不是所有的扩展报头都需要被转发路由设备查看和处理的。路由设备转发时根据基本报头中 Next Header 值来决定是否要处理扩展头。

除了目的选项扩展报头出现两次（一次在路由扩展报头之前，另一次在上层扩展报头之前），其余扩展报头只出现一次。

## 4.4.3 ICMPv6

ICMPv6（Internet Control Message Protocol for the Internet Protocol Version 6）是 IPv6 的基础协议之一，具有差错报文和信息报文两种，用于 IPv6 节点报告报文处理过程中的错误和信息。ICMPv6 报文的报文格式如图 4-4 所示。

**图 4-4** ICMPv6 报文格式




报文中各个字段的解释如下：

- Type 字段表明消息的类型，0 至 127 表示差错报文类型，128 至 255 为消息报文类型。
- Code 字段表示此消息类型细分的类型。

- Checksum 表示 ICMPv6 报文的校验和。

### ICMPv6 错误报文的分类

- 目的不可达错误报文

  在 IPv6 节点转发 IPv6 报文过程中，发现目的地址不可达时，就会向发送报文的源节点发送 ICMPv6 目的不可达错误报文。同时报文中会携带引起该错误报文的具体原因。目的不可达错误报文又细分为以下几种：

  - 没有到目的地的路由
  - 地址不可达
  - 端口不可达

- 数据包过大错误报文

  在 IPv6 节点转发 IPv6 报文过程中，发现报文超过出接口的链路 MTU 时，则向发送报文的源节点发送 ICMPv6 数据包过大错误报文，其中携带出接口的链路 MTU 值。数据包过大错误报文是 Path MTU 发现机制的基础。

- 时间超时错误报文

  在 IPv6 报文收发过程中，当设备收到 Hop Limit 值等于 0 的数据包，或者当设备将 HopLimit 值减为 0 时，会向报文的源节点发送 ICMPv6 超时错误报文。对于分段重组报文的操作，如果超过定时时间，也会产生一个 ICMPv6 超时报文。

- 参数错误报文

  当目的节点收到一个 IPv6 报文时，会对报文进行有效性检查，如果发现以下问题会向报文的源节点回应一个 ICMPv6 参数错误差错报文。

  - IPv6 基本头或扩展头的某个域有错误
  - IPv6 基本头或扩展头的 NextHeader 值不可识别
  - 扩展头中出现未知的选项

### ICMPv6 信息报文的分类

请求信息（Echo Request）和应答信息（Echo Reply）。可以利用 ICMPv6 报文实现网络故障诊断、PMTU 发现和邻居发现等功能。在两节点的互通性检测中，收到 Echo Request 报文的节点向源节点回应 Echo Reply 报文，实现两节点间报文的收发。

## 4.4.4 邻居发现

邻居发现 ND（Neighbor Discovery）是确定邻居节点之间关系的一组消息和进程。邻居发现协议替代了 IPv4 的 ARP（Address Resolution Protocol）、ICMP 路由器发现（Router Discovery）和 ICMP 重定向（Redirect）消息，并提供了其他功能。

对于一个节点而言，当其配置一个 IPv6 地址之后，首先会确定此地址是否可用、不冲突。当一个节点是主机时，路由器需要通知主机向特定目的地址转发报文的理想下一跳地址；当一个节点是路由器时，需要发布自己的地址、地址前缀和其他配置参数以指导主机进行参数配置。在 IPv6 报文转发过程中，节点需要确定邻居节点的链路层地址和其可达性。IPv6 邻居发现机制提供了 5 种不同类型的 ICMPv6 报文。

- 路由器请求报文 RS（Router Solicitation）：主机启动后，通过 RS 报文向路由设备发出请求，路由设备则会以 RA 报文响应。
- 路由器通告报文 RA（Router Advertisement）：路由设备周期性的发布 RA 报文，其中包括前缀和一些标志位的信息。

- 邻居请求报文 NS（Neighbor Solicitation）：IPv6 节点通过 NS 报文可以得到邻居的链路层地址，检查邻居是否可达，也可以进行重复地址检测。
- 邻居通告报文 NA（Neighbor Advertisement）：NA 报文是 IPv6 节点对 NS 报文的响应，同时 IPv6 节点在链路层变化时也可以主动发送 NA 报文。
- 重定向报文（Redirect）：路由设备发现报文的入接口和出接口相同时，可以通过重定向报文通知主机选择另外一个更好的下一跳地址。

  说明

  AR150/200 暂不支持重定向 IPv6 报文。

IPv6 邻居发现协议主要包括以下功能：

## 地址冲突检测功能

地址冲突检测 DAD（Duplicate address detect）是确定 IPv6 地址是否可用的一种探测机制。具体执行过程如下：

1. 当一个节点配置了 IPv6 地址，为了查看该地址是否被其他邻居节点所使用，会即时发送邻居请求报文来确定其可用性。
2. 当其他邻居节点收到该报文后会查找本地的 IPv6 地址中是否存在相同的 IPv6 地址，若存在会回应一个邻居通告报文给源节点，并携带此 IPv6 地址信息。
3. 源节点收到邻居的回应报文则认为该 IPv6 地址已被邻居使用。反之，如果源节点发出的邻居请求报文没有收到相应的回应报文，则表示配置的 IPv6 地址是可用的。

## 邻居发现功能

邻居发现功能和 IPv4 中的 ARP 功能类似，主要实现对邻居地址的解析和邻居可达性的探测，依赖于邻居请求和邻居通告报文完成。

当一个节点需要得到同一本地链路上另外一个节点的链路层地址时，就会发送 ICMPv6 类型为 135 的邻居请求报文。此报文类似于 IPv4 中的 ARP 请求报文，不过使用组播地址而不使用广播地址，只有被请求节点的最后 24 比特和此组播地址相同的节点才会收到此报文，减少了广播风暴的可能。目的节点在响应报文中填充其链路层地址。

邻居请求报文也用来在邻居的链路层地址已知时，验证邻居的可达性。IPv6 邻居通告报文是对 IPv6 邻居请求报文的响应。收到邻居请求报文后，目的节点通过在本地链路上发送 ICMPv6 类型为 136 的邻居通告报文进行响应。收到邻居通告后，源节点和目的节点可以进行通信。当一个节点的本地链路上的链路层地址改变时也会主动发送邻居通告报文。

## 路由器发现功能

路由器发现功能用来定位邻居路由设备，同时学习和地址自动配置有关的前缀和配置参数。IPv6 路由发现由下面两种机制实现：

- 路由器请求

  当主机没有配置单播地址时（例如系统刚启动），就会发送路由器请求报文 RS。路由器请求报文有助于主机迅速进行自动配置而不必等待 IPv6 路由设备的周期性路由器通告报文。IPv6 路由器请求也是 ICMPv6 报文，类型为 133。
- 路由器通告

  每个 IPv6 路由设备的接口在配置了 IPv6 RA 去抑制的前提下会周期发送路由器通告报文。在本地链路上收到 IPv6 节点的路由器请求报文后，路由设备也会回应路由器通告报文。IPv6 路由器通告报文发送到所有节点多播地址（FF02::1）或发送

路由器请求报文节点的 IPv6 单播地址。路由器通告为 ICMPv6 报文，类型为 134，包含以下内容：

- 是否使用地址自动配置
- 标记支持的自动配置类型（无状态或有状态自动配置）
- 一个或多个本地链路前缀（本地链路上的节点可以使用这些前缀完成地址自动配置）
- 通告的本地链路前缀的生存期
- 发送路由器通告的路由设备是否可作为缺省路由设备，如果可以，还包括此路由设备可作为缺省路由设备的时间（用秒表示）
- 和主机相关的其它信息，如跳数限制、主机发起的报文可以使用的最大 MTU

本地链路上的 IPv6 节点接收路由器通告报文，并用其中的信息得到更新的缺省路由设备、前缀列表以及其它配置。

## 地址自动配置功能

通过使用路由器通告报文和针对每一前缀的标记，路由设备可以通知主机如何进行地址自动配置。例如，路由设备可以指定主机是使用有状态（DHCPv6）地址配置还是无状态地址自动配置进行地址配置。

对于无状态地址自动配置而言，当主机收到路由器通告报文后，使用其中的前缀信息和本地接口 ID 自动形成 IPv6 地址，同时还可以根据其中的默认路由设备信息设置默认路由设备。

## IPv6 安全邻居发现功能

IPv6 邻居发现协议（NDP，Neighbor Discovery Protocol）用来保证本链路内邻居的可达性，因此非常有必要保护 NDP 的安全。IPSec 方法可以在一定程度上保护 NDP，但是这种方法需要大量且复杂的手工配置。此时可以通过简单配置 IPv6 安全邻居发现（SEND，Security Neighbor Discovery）特性，实现对 IPv6 邻居发现协议的保护。

说明

AR150/200 暂不支持 IPv6 安全邻居发现功能。

SEND 特性用来解决在 NDP 中涉及的安全问题，如：

- 重定向攻击：NS/NA 欺骗、恶意的最后一跳路由器、虚假的重定向报文、重放攻击。

  攻击节点可以使用携带不同源/目的链路层地址选项的 NS 报文，通过 NS/NA 欺骗，使合法节点的报文发往其他的链路层地址达到攻击的目的。

- 拒绝服务攻击（DoS，Denial of Service）：邻居不可达探测（NUD：Neighbor Unreachibility Detection）失败、DAD 攻击、虚假的地址配置前缀、参数欺骗。

  攻击者持续发送虚假的 NA 响应 NUD 的 NS，主机经过几次重试失败后就会删除被攻击者的邻居表项记录，造成被攻击者无法通信。攻击者还可以通过响应所有的 DAD 过程，通告已使用了被攻击者请求的地址，造成被攻击者因获取不到 IP 地址而无法正常运行。

为了解决上面的安全问题，SEND 中引入两种新的选项：CGA 选项（Cryptographically Generated Addresses）和 RSA 选项（Rivest Shamir Adleman）。

- CGA 是一种新的地址自动生成机制，可以用来验证 ND 报文的发送者对报文源地址拥有权的合法性（此地址指 ND 报文的源地址）。

- RSA 是对 ND 报文的数字签名，用来验证报文的完整性和发送者的真实性。

SEND 特性同时还定义了 ND 报文中的两种选项，用以解决 NDP 的安全问题：

- 随机数（Nonce）选项：用来在请求和回应交互中防止重放攻击，比如在 NS 和 NA 报文的交互中，NS 报文中携带 Nonce 选项，回应的 NA 报文中也携带此选项，发送者根据收到的选项判断是否为合法的回应报文。
- 时间戳（Timestamp）选项：用来保护非请求的通告报文和重定向报文。接收者应确保每个收到的报文其时间戳都比上一个收到的报文要新。

当接口需要拒绝接收非安全的 ND 报文时，可以配置 IPv6 安全邻居发现功能。如果满足以下条件中的任何一条，即为非安全的 ND 报文：

- 接收到的 ND 报文没有携带 CGA 和 RSA 选项，即发送该报文的接口没有配置 CGA 地址。
- 接收到的 ND 报文的密钥长度超出本接口可以接受的长度范围。
- 接收 ND 报文的速率超出系统接受的速率范围。
- 接收到 ND 报文的时间与发送 ND 报文的时间差超出本接口可以接受的时间差范围。

### 默认路由器优先级和路由信息

在邻居发现协议的 RA 报文中，定义了默认路由器优先级和路由信息两个字段，帮助主机在发送报文时选择合适的转发路由器。

当主机所在的链路中存在多个路由器时，主机需要根据报文的目的地址选择转发路由器。在这种情况下，路由器通过发布默认路由器优先级和特定路由信息给主机，提高主机根据不同的目的地选择合适的转发路由器的能力。

主机收到包含路由信息的 RA 报文后，会更新自己的路由表。当主机向其他设备发送报文时，通过查询该列表的路由信息，选择合适的路由发送报文。

主机收到包含默认路由器优先级信息的 RA 报文后，会更新自己的默认路由器列表。当主机向其他设备发送报文时，如果没有路由可选，则首先查询该列表，然后选择本链路内优先级最高的路由器发送报文；如果该路由器故障，主机根据优先级从高到低的顺序，依次选择其他路由器。

## 4.4.5 Path MTU

### 网络上的 MTU

由于 IPv6 报文在传输过程中不允许在中间节点分片转发，所以在转发过程中经常会出现报文长度大于路径 IPv6 MTU 的情形，这就需要源节点不断的进行重传，降低了传输的效率，如果在源节点使用最小链接 IPv6 MTU（1280）作为分片的最大长度，在大多数情况下，路径的 IPv6 MTU 是大于最小链接的 IPv6 MTU 的，一个节点发出的分片远小于路径 IPv6 MTU，这是对网络资源的一种浪费，为了解决这个问题，提出了路径 MTU 发现协议。

### Path MTU 的工作原理

Path MTU（以下简称 PMTU），是确定从源端到目的端路径上合适的 IPv6 MTU 值的一种机制。PMTU 发现协议描述了一种动态发现任意路径的 PMTU 的方法。当一个 IPv6 节点发送大量数据到另一节点时，数据通过一系列 IPv6 分片传送。当这些分片具有从

源节点到信宿节点能够成功传送所允许的最大长度时，我们认为它达到理想状态，这个分片长度被称为路径 MTU。

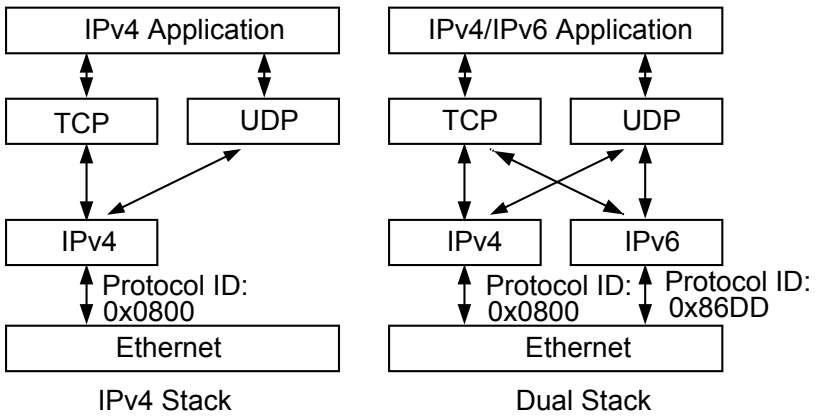
一个源节点开始会假设一个路径的 PMTU 是路径中第一跳的已知的 IPv6 MTU，如果从那个路径发出的报文太大以至于不能沿着路径转发，中间节点将丢弃此报文并返回一个 ICMPv6 数据过大差错报文给源节点，根据数据过大消息中的 IPv6 MTU 值来设置此路径的 PMTU 值。

当节点学习到的 PMTU 值小于或者等于实际的 PMTU 时，PMTU 的发现过程结束。注意在 PMTU 发现过程结束之前，可能会出现反复发送报文和收到报文太大消息，这是因为可能会不断发现更远的路径链路有更小的 IPv6 MTU。

## 4.4.6 双协议栈

对于 IPv6 节点来说，兼容 IPv4 的最直接有效的办法就是保留一个完整的 IPv4 协议栈，这样的节点即为双协议栈节点。单协议栈和双协议栈结构示例如图 4-5 所示。

**图 4-5 单协议栈与双协议栈结构（以太网）**




双协议栈具有以下特点：

- 多种链路协议支持双协议栈

  多种链路协议（如以太网）支持双协议栈。图中的链路层是以太网，在以太网帧上，如果协议 ID 字段的值为 0x0800，表示网络层收到的是 IPv4 报文，如果为 0x86DD，表示网络层是 IPv6 报文。

- 多种应用支持双协议栈

  多种应用（如 DNS/FTP/Telnet 等）支持双协议栈。上层应用（如 DNS）可以选用 TCP 或 UDP 作为传输层的协议，但优先选择 IPv6 协议栈，而不是 IPv4 协议栈作为网络层协议。

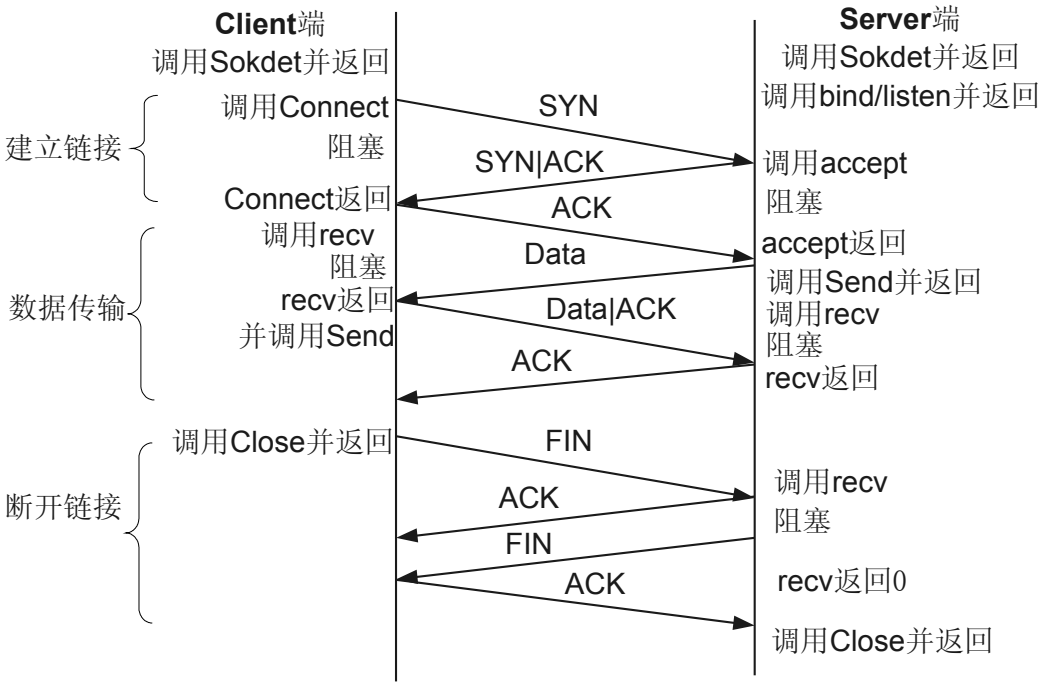
## 4.4.7 TCP6

TCP6 提供了在两个端点的进程间建立虚电路的机制，一个 TCP6 虚连接如同在系统间承载数据的全双工电路。由于 TCP6 中提供了进程间数据的可靠传输，因此被称为可靠协议，它还提供了根据当前网络状态来优化传输性能的机制。在所有数据均可收到和确认的情况下，传输速率可以逐渐增加。延时将导致发送主机在收到进一步的确认前降低发送速率。

TCP6 通常用于交互式应用，如 WEB 之类，某些数据接收差错将影响正常的工作能力。TCP6 使用了“三次握手”机制来建立虚电路，所有的虚电路都需使用“四次握手”拆除。这种连接方式可以提供多种校验和及其他可靠性功能，但是增加了使用 TCP6 的开销并导致其效率低于UDP6。

如图 4-6 表示了 TCP6 连接建立和拆除的过程。

**图 4-6** TCP6 连接建立和拆除过程示意图




## 4.4.8 UDP6

UDP6 是用来在互连网络环境中提供包交换的计算机通信协议。有如下特点：

- 只使用源和目的信息，主要用于简单的请求/响应式结构。
- 不可靠，即没有任何控制能确定 UDP6 数据报是否已被接收。
- 无连接，即在主机间传输数据时，不需要任何类型的虚电路。

UDP6 的无连接特性使得 UDP6 可以向广播地址发送数据；而 TCP6 则不同，它要求特定的源地址和目的地址。

## 4.4.9 RawIP6

RawIP6 较为简单，只填充 IPv6 首部的有限几个字段，允许应用进程提供自己的 IPv6 首部。

RawIP6 类似于 UDP6：

- 不可靠，即没有任何控制能确定 RawIP6 数据报是否已被接收。
- 无连接，即在主机间传输数据时，不需要任何类型的虚电路。

RawIP6 相比 UDP6 的区别在于，RawIP6 允许应用程序直接通过 Socket 接口操作 IP 层。对于许多需要跟下层直接交互的应用来说，非常方便。

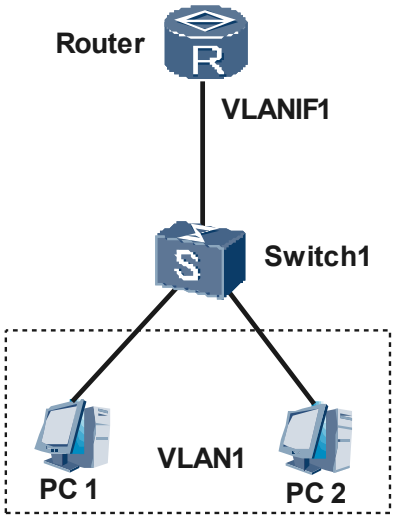
# 4.5 应用

## VLAN 内 ND 代理

在 IPv6 网络中，两个用户属于相同的 VLAN，但 VLAN 内配置了用户隔离，此时如果用户间进行互通，需要在 VLANIF 接口上使能 VLAN 内 ND 代理功能。

如图 4-7 所示，PC1 和 PC2 是属于 VLAN1 内的两个用户。在中间交换机 S1 配置了 VLAN 内不同接口彼此隔离，因此 PC1 和 PC2 不能直接互通。通过在 Router 的 VLANIF1 接口上配置 VLAN 内 ND 代理功能，实现 PC1 和 PC2 的互通。

**图 4-7** VLAN 内 ND 代理典型组网图




PC1 与 PC2 互通时，PC1 先发送一个 NS 请求报文请求 PC2 的 MAC 地址，由于中间 S1 设备配置了隔离，所以 NS 请求报文无法发送至 PC2，此时路由设备把 NS 报文转发至 PC2，并把 NS 报文携带的 MAC 地址改为路由设备 VLANIF1 的 MAC 地址。PC2 回应 NA 报文给 PC1 时，路由设备接收到此 NA 回应报文后会生成一条 PC2 的 ND 表项，并生成对应的路由表项，同时会把此 NA 报文的 MAC 地址改变为本路由设备的 MAC 地址，然后转发至 PC1。这样 PC1 学习到 PC2 的 MAC 为路由设备的 MAC。

PC1 根据学习到的 ND 表项封装报文，把报文发送到 Router，并由 Router 根据上面学到的主机 PC2 的路由把报文转发到 PC2。

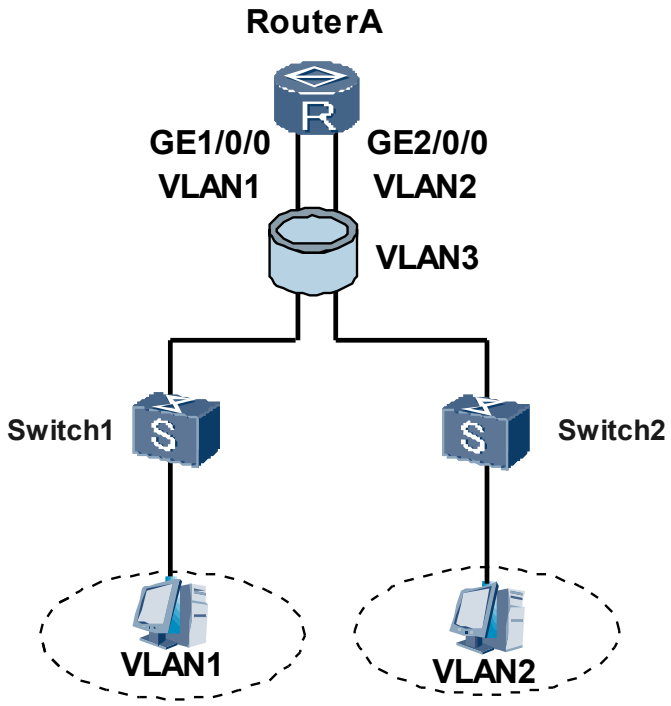
## VLAN 间 ND 代理

如果两个用户属于不同的 VLAN，用户间要进行互通，需要在 Aggregate-VLAN（聚合 VLAN）对应的 VLANIF 接口上启动 VLAN 间 ND 代理功能，实现 Access-VLAN 间用户的互通。

如图 4-8 所示，PC1 和 PC2 分别通过交换机 S1 和 S2 是连接到路由设备的 VLAN1（Access-VLAN）和 VLAN2（Access-VLAN），VLAN1 与 VLAN2 属于 VLAN3

（Aggregate-VLAN）。通过在 Router 的 VLANIF3 接口上配置 VLAN 间 ND 代理功能，实现 PC1 和 PC2 的互通。

**图 4-8** VLAN 间 ND 代理典型组网图




PC1 与 PC2 互通时，PC1 先发送一个 NS 请求报文请求 PC2 的 MAC 地址，此时路由设备把 NS 报文转发至 PC2，并把 NS 报文携带的 MAC 地址改为路由设备 VLANIF1 的 MAC 地址，PC2 回应 NA 报文给 PC1 时，路由设备接收到此 NA 回应报文后会生成一条 PC2 的 ND 表项，并生成对应的路由表项，同时会把此 NA 报文的 MAC 地址改变为本路由设备的 MAC 地址，然后转发至 PC1。这样 PC1 学习到 PC2 的 IPv6 对应的 MAC 为路由设备的 MAC。

PC1 根据学习到的 ND 表项封装报文，把报文发送到 Router，并由 Router 根据上面学到的主机 PC2 的路由把报文转发到 PC2。

# 4.6 术语与缩略语

## 术语

| 术语 | 解释 |
| --- | --- |
| IPv6 | Internet Protocol Version 6，下一代网际协议 |
| ND | Neighbor Discovery，邻居发现，在 IPv6 报文转发过程中，用于地址冲突检测、邻居地址解析、确定邻居可达性，以及进行主机地址配置的一组协议和进程。由不同的 ICMPv6 报文实现路由器发现和邻居发现功能。 |

| 术语 | 解释 |
| --- | --- |
| ICMPv6 | Internet Control Management Protocol Version 6，Internet 互联网控制报文协议第 6 版，是 IPv6 的基础协议之一，具有差错报文和信息报文两种，用于 IPv6 结点报告报文处理过程中的错误和信息。 |
| PMTU | Path MTU，路径 MTU，利用 ICMPv6 数据包过大差错报文确定路径支持的最大传输单元的方法。 |

## 缩略语

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| IPv6 | Internet Protocol Version 6 | 网际协议第 6 版 |
| ICMPv6 | Internet Control Management Protocol Version 6 | Internet 互联网控制报文协议第 6 版 |
| ND | Neighbor Discovery | 邻居发现 |
| RS | Router Solicitation | 路由器请求 |
| RA | Router Advertisement | 路由器通告 |
| NS | Neighbor Solicitation | 邻居请求 |
| NA | Neighbor Advertisement | 邻居通告 |
| ARP | Address Resolution Protocol | 地址解析协议 |
| PMTU | Path MTU | 路径 MTU |
| IPng | IP Next Generation | 网络层协议的第二代标准协议 |
| TCP6 | Transmission Control Protocol 6 | 传输控制协议 6 |
| UDP6 | User Datagram Protocol 6 | 用户数据报协议 6 |
| RawIP6 | Raw IP6 | 原始 IP6 |

# 5 DNS

## 关于本章

# 5.1 介绍

## 定义

TCP/IP 提供了通过 IP 地址来确定设备的功能，但对用户来讲，记住某台设备的 IP 地址是相当困难的，因此专门设计了一种字符串形式的主机命名机制，这些主机名与 IP 地址一一对应。在 IP 地址与主机名之间需要有一种转换和查询机制，提供这种机制的系统就是域名系统 DNS（Domain Name System）。

## 目的

域名系统 DNS 使用一种有层次的命名方式，为网上的设备指定一个有意义的名字，并且在网络上设置域名解析服务器，建立域名与 IP 地址的对应关系。这样用户就可以使用便于记忆的、有意义的域名，而不必去记忆复杂的 IP 地址。

# 5.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC1034 | DOMAIN NAMES - CONCEPTS AND FACILITIES | - |
| RFC1035 | DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION | - |

# 5.3 可获得性

## 涉及网元

无需其它网元的配合。

## License 支持

无需获得 License 许可，均可获得该特性的服务。

## 版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

### 特性依赖

不依赖其他特性。

### 硬件要求

对硬件无特殊要求。

## 5.4 原理描述

域名解析分为动态解析和静态解析，二者可以相辅相成，可以配合使用。在解析域名时，先采用静态解析（通过查找静态域名解析表）的方法，如果静态解析不成功，再采用动态解析的方法。将一些常用的域名放入静态域名解析表中，可以提高域名解析效率。

### 5.4.1 静态 DNS

如果用户使用域名访问其他设备的次数很少，或者没有可用的 DNS 服务器时，需要配置静态 DNS。配置静态 DNS 需要网络管理员知道域名与 IP 地址的对应关系，且在域名与 IP 地址的对应关系变化时，需要手动修改 DNS 表项。

静态域名解析通过静态域名解析表进行，即手动建立域名和 IP 地址之间的对应关系表，将一些常用的域名放入表中。当客户机需要域名所对应的 IP 地址时，首先到静态域名解析表中查找指定的域名，从而获得所对应的 IP 地址，提高域名解析的效率。

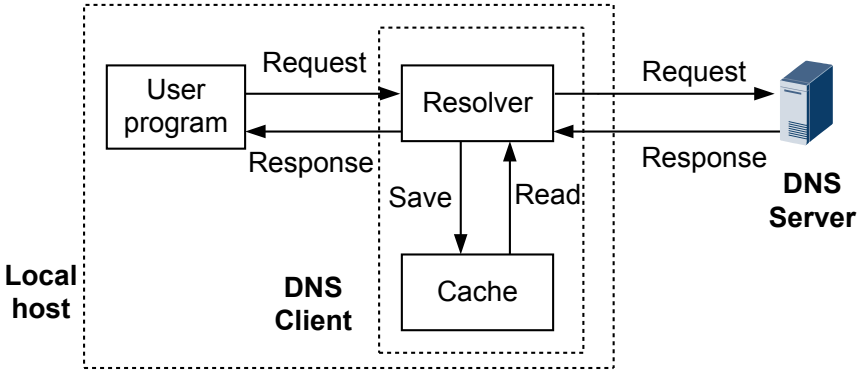
### 5.4.2 动态 DNS

动态域名解析有专用的 DNS 服务器，负责接受 DNS 客户端提出的域名解析请求。DNS 服务器首先在本机数据库内部解析，如果判断不属于本域范围之内，就将请求交给上一级的 DNS 服务器，直到完成解析，解析的结果或者为 IP 地址，或者域名不存在，并将解析的结果反馈给 DNS 客户端。

用户程序（例如 Ping、Tracert）对 DNS 服务器（DNS Server）的访问是通过 DNS 客户端（DNS 客户端）的一个地址解析器（Resolver）完成的。

用户程序（例如 Ping、Tracert）、解析器和 DNS 服务器以及解析器上的缓存区关系如图 5-1 所示。

**图 5-1** 动态 DNS



其中解析器和缓存区集成在一起构成 DNS 客户端，它的作用是接受用户程序的 DNS 咨询，并对其做出反应。一般来说，用户程序（例如 Ping、Tracert）、缓存区和解析器是在同一台主机上，DNS 服务器在另外的主机上。

## 动态 DNS 的工作过程

1. 用户程序（例如 Ping、Tracert）首先向 DNS 客户端发出请求；
2. DNS 客户端收到请求后，首先查询本机数据库/缓存，如果没有发现所要查找的映射项，就向 DNS 服务器发送查询报文；
3. DNS 服务器收到查询报文后，首先判断请求的域名是否处于自己被授权管理的子域里，再根据不同的判断结果，向 DNS 客户端发送相应的响应报文；
4. DNS 客户端收到响应后，解析 DNS 服务器发回来的响应报文，并根据响应报文的内容决定下一步的操作。

## 域名后缀列表功能

动态域名解析支持域名后缀列表功能，用户可以预先设置一些域名后缀，在域名解析的时候，用户只需要输入域名的部分字段，系统会自动将输入的域名加上不同的后缀进行解析。

## 域名解析方式

动态域名解析需要专用的域名解析服务器，该服务器运行域名解析服务器程序，提供从域名到 IP 地址的映射关系，负责处理客户提出的域名解析请求。

域名解析服务器接收到客户端提出的域名解析请求后，首先判断请求的域名是否处于自己被授权管理的子域里。如果是，就查询数据库，把域名转换为 IP 地址，并将转换结果发送给客户端。如果域名解析服务器不能解析出域名，它就根据客户在查询报文中所指明的解析方式（递归解析或者迭代解析）来进行下一步操作。

有以下两种域名解析方式：

- 递归解析

  域名解析服务器和其他能解析该域名的服务器联系，并将查询结果即域名所对应的 IP 地址返回给客户端。

- 迭代解析

  若该域名解析服务器不能提供解析结果，会在给客户端的响应报文中指明客户端应联系的下一个域名解析服务器。客户端会向指明的下一个域名解析服务器再次发出查询请求。

## 查询类型

目前，设备支持 DNS 客户端功能，支持 A 类查询、PTR 类、SRV 查询和 NAPTR 查询。

- A 类查询是最常用的查询类型，用于获取域名对应的 IP 地址。例如在 ping 和 tracert 的时候，可以 ping 或 tracert 一个域名，此时 ping 或 tracert 作为用户程序会向系统中 DNS 客户端查询该域名对应的 IP 地址。如果系统中没有该域名对应的 IP 地址信息，DNS 客户端就会向 DNS 服务器发起 A 类查询，获取该域名对应的 IP 地址，完成 ping 和 tracert 的功能。
- PTR 类查询用于获取 IP 地址对应的域名。

- SRV 查询根据某种协议查找应用该协议的服务器的相关信息，包括服务器域名、端口号等信息。
- NAPTR 查询根据应用服务器的域名来获取该服务器的相关信息，包括下一步查询的名字、传输协议等信息。

# 5.5 应用

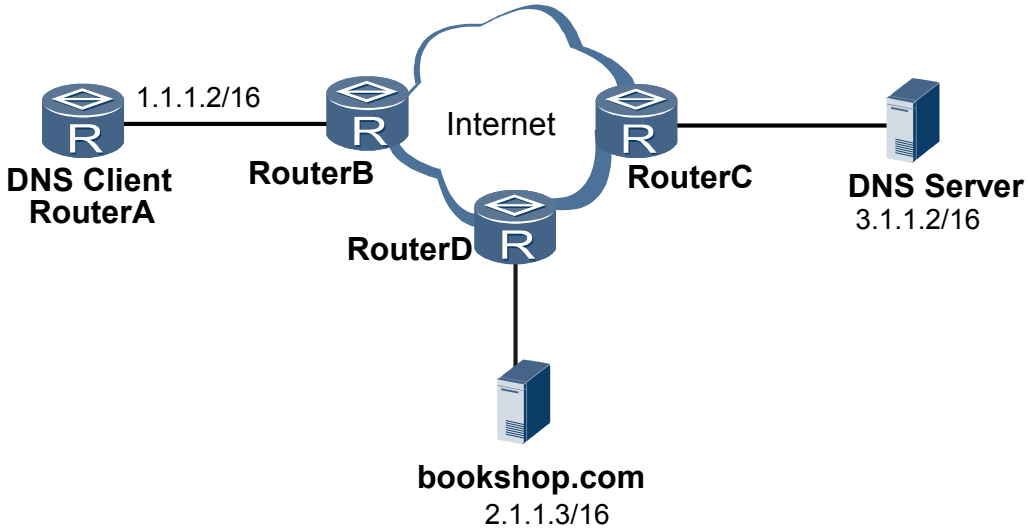
由于 Internet 协议（IP）地址结构不便于记忆（比如点分式表示的 202.112.131.109），所以大多数组织采用缩写词或有意义的名字（称为域名，如 www.sina.com.cn）来表示地址，而不是使用 IP 地址。但是，如何让非 IP 标识的域名映射为 IP 地址呢？IP 地址与其域名之间的映射是依靠解析器及域名服务器来完成的。DNS 客户端主要完成解析器的功能，它的主要功能是完成 IP 地址和主机域名之间的转换。

如果用户使用域名访问其他设备的次数很少，或者没有可用的 DNS 服务器时，需要配置静态 DNS。配置静态 DNS 需要网络管理员知道域名与 IP 地址的对应关系，且在域名与 IP 地址的对应关系变化时，需要手动修改 DNS 表项。

如果用户需要使用域名访问很多的设备，且有可用的 DNS 服务器，此时可配置动态 DNS。动态 DNS 需要有 DNS 服务器的支持。

如图 5-2 所示，RouterA 作为 DNS 客户端与 DNS 服务器配合，使 RouterA 能够通过域名（www.huaweibookshop.com）访问 IP 地址为 2.1.1.3/16 的主机。

**图 5-2** 配置 DNS 客户端组网图




# 5.6 术语与缩略语

## 术语

| 术语 | 解释 |
| --- | --- |
| DNS Server | DNS 域名服务器。能在网络上给客户端提供域名解析服务的设备。 |

| 术语 | 解释 |
| --- | --- |
| DNS Client | DNS 客户端。向 DNS 域名服务器发出请求并等待响应的设备。 |

## 缩略语

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| DNS | Domain Name System | 域名系统 |

# 6 DHCP

## 关于本章

# 6.1 介绍

## 定义

DHCP（Dynamic Host Configuration Protocol）是一种用于集中对用户主机的配置信息进行动态管理和配置的技术。

DHCP 采用客户端/服务器通信模式，由客户端向服务器提出配置申请（包括 IP 地址、子网掩码、缺省网关等参数），服务器根据策略返回相应配置信息。

## 目的

随着网络规模的扩大和网络复杂度的提高，网络配置越来越复杂，计算机位置变化（如便携机或无线网络）和计算机数量超过可分配的 IP 地址，造成 IP 地址变化频繁以及 IP 地址不足的情况。为了动态合理地分配 IP 地址给主机使用，需要用到 DHCP。

DHCP 协议是在 BOOTP（Bootstrap Protocol）协议基础上发展而来，但 BOOTP 运行在相对静态（每台主机都有固定的网络连接）的环境中，管理员为每台主机配置专门的 BOOTP 参数文件，该文件会在相当长的时间内保持不变。而 DHCP 从两方面对 BOOTP 进行了扩展：

- DHCP 允许计算机快速、动态地获取 IP 地址，而不是静态为每台主机指定地址。
- DHCP 加入了 IP 地址的动态管理功能和获取附加配置选项的功能，可使计算机获取它所需要的所有配置信息。

DHCP 技术保证了 IP 地址的合理分配问题，从而避免了 IP 地址的浪费，提高了整网的 IP 地址使用率。

## 受益

企业受益。

DHCP 技术实现了用户 IP 地址和配置信息的动态分配和集中管理，使企业可以快速、动态地为企业用户分配和管理地址，避免繁琐的手工配置，快速适应网络的变化。

# 6.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 | 备注 |
| --- | --- | --- |
| RFC1534 | Interoperation Between DHCP and BOOTP | - |
| RFC2131 | Dynamic Host Configuration Protocol | - |
| RFC2132 | DHCP Options and BOOTP Vendor Extensions | - |
| RFC3046 | DHCP Relay Agent Information Option | - |

# 6.3 可获得性

## 涉及网元

DHCP 服务器、DHCP 中继代理、DHCP 客户端。

## 版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

## 特性依赖

不依赖其他特性。

## 硬件要求

对硬件无特殊要求。

## 性能规格

| 规格项 | AR150/200 | AR1200 | AR2200 | AR3200 |
| --- | --- | --- | --- | --- |
| 每个地址池下最多可分配的地址数目 | 256 | 512 | 1024 | 2048 |
| 可配置为客户端的接口的最大数目 | 32 | 32 | 32 | 32 |
| 全局地址池的最大数目 | 16 | 64 | 128 | 128 |
| 接口地址池的最大数目 | 16 | 64 | 128 | 128 |
| 每个全局地址池最多可配置的出口网关数目 | 8 | 8 | 8 | 8 |
| 每个地址池下最多配置的 DNS 服务器数目 | 8 | 8 | 8 | 8 |
| 每个地址池下最多配置的 NetBIOS 服务器数目 | 8 | 8 | 8 | 8 |
| 地址池下 IP/MAC 绑定的最大数目 | 256 | 512 | 1024 | 2048 |

# 6.4 原理描述

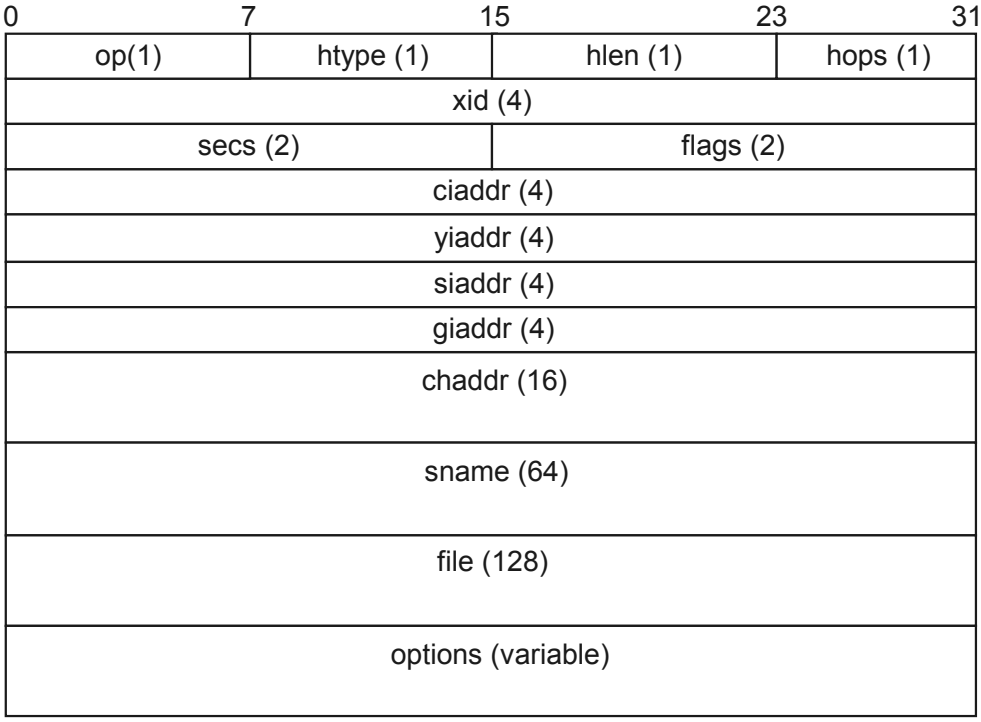
## 6.4.1 DHCP 报文

DHCP 报文分为 8 种类型，DHCP 服务器和客户端之间通过这 8 种类型的报文进行通信。

- DHCP DISCOVER：这是 DHCP 客户端首次登录网络时进行 DHCP 过程的第一个报文，用来寻找 DHCP 服务器。
- DHCP OFFER：DHCP 服务器用来响应 DHCP DISCOVER 报文，此报文携带了各种配置信息。
- DHCP REQUEST：此报文用于以下三种用途。
  - 客户端初始化后，发送广播的 DHCP REQUEST 报文来回应服务器的 DHCP OFFER 报文。
  - 客户端重启初始化后，发送广播的 DHCP REQUEST 报文来确认先前被分配的 IP 地址等配置信息。
  - 当客户端已经和某个 IP 地址绑定后，发送 DHCP REQUEST 报文来延长 IP 地址的租期。
- DHCP ACK：服务器对客户端的 DHCP REQUEST 报文的确认响应报文，客户端收到此报文后，才真正获得了 IP 地址和相关的配置信息。
- DHCP NAK：服务器对客户端的 DHCP REQUEST 报文的拒绝响应报文，比如服务器对客户端分配的 IP 地址已超过使用租借期限或者客户端移到了另一个新的网络。
- DHCP DECLINE：当客户端发现服务器分配给它的 IP 地址发生冲突时会通过发送此报文来通知服务器，并且会重新向服务器申请地址。
- DHCP RELEASE：客户端可通过发送此报文主动释放服务器分配给它的 IP 地址，当服务器收到此报文后，可将这个 IP 地址分配给其它的客户端。
- DHCP INFORM：客户端已经获得了 IP 地址，发送此报文的目的是为了从服务器获得其他的一些网络配置信息，比如网关地址、DNS 服务器地址等。

以上 8 种类型报文的格式相同，只是某些字段的取值不同。DHCP 报文格式基于 BOOTP 的报文格式，具体格式如图 6-1 所示（括号中的数字表示该字段所占的字节）。

**图 6-1** DHCP 报文格式




各字段的解释如下：

- op：报文的操作类型，分为请求报文和响应报文，“1”为请求报文；“2”为响应报文。具体的报文类型在 option 字段中标识。
- htype、hlen：DHCP 客户端的硬件地址类型及长度。
- hops：DHCP 报文经过的 DHCP 中继的数目。DHCP 请求报文每经过一个 DHCP 中继，该字段就会增加 1。
- xid：客户端发起一次请求时选择的随机数，用来标识一次地址请求过程。
- secs：DHCP 客户端开始 DHCP 请求后所经过的时间。目前没有使用，固定为 0。
- flags：第一个比特为广播响应标识位，用来标识 DHCP 服务器响应报文是采用单播还是广播方式发送，0 表示采用单播方式，1 表示采用广播方式。其余比特保留不用。
- ciaddr：DHCP 客户端的 IP 地址。
- yiaddr：DHCP 服务器分配给客户端的 IP 地址。
- siaddr：服务器 IP 地址。
- giaddr：DHCP 客户端发出的请求报文所经过的第一个 DHCP 中继的 IP 地址。
- chaddr：DHCP 客户端的硬件地址。
- sname：DHCP 客户端获取 IP 地址等信息的服务器名称。
- file：DHCP 服务器为 DHCP 客户端指定的启动配置文件名称及路径信息。
- options：可选变长选项字段，包含报文的类型、有效租期、DNS 服务器的 IP 地址、WINS 服务器的 IP 地址等配置信息。

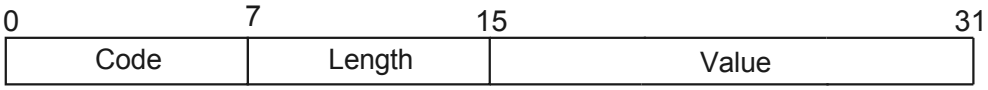
## 6.4.2 DHCP 选项

DHCP 利用 Option 字段传递控制信息和网络配置参数，实现地址的动态分配，为客户端提供更加丰富的网络配置信息。

### DHCP 选项格式

DHCP 选项的格式如图 6-2 所示。

**图 6-2** DHCP 选项格式




### 常见 DHCP 选项

常见的 DHCP 选项有：

- Option3：路由器选项，用来指定为客户端分配的网关地址。
- Option6：DNS 服务器选项，用来指定为客户端分配的 DNS 服务器地址。
- Option51：IP 地址租约选项。
- Option53：DHCP 消息类型选项，标识 DHCP 消息的类型。
- Option55：请求参数列表选项。客户端利用该选项指明需要从服务器获取哪些网络配置参数。该选项内容为客户端请求的参数对应的选项值。
- Option66：TFTP 服务器名选项，用来指定为客户端分配的 TFTP 服务器的域名。
- Option67：启动文件名选项，用来指定为客户端分配的启动文件名。
- Option150：TFTP 服务器地址选项，用来指定为客户端分配的 TFTP 服务器的地址。
- Option121：无分类路由选项。该选项中包含一组无分类静态路由（即目的地址的掩码为任意值，可以通过掩码来划分子网），客户端收到该选项后，将在路由表中添加这些静态路由。
- Option33：静态路由选项。该选项中包含一组有分类静态路由（即目的地址的掩码固定为自然掩码，不能划分子网），客户端收到该选项后，将在路由表中添加这些静态路由。如果存在 Option121，则忽略该选项。

更多 DHCP 选项的介绍，请查看 RFC2132。

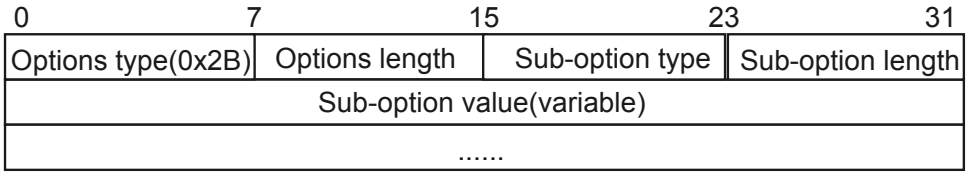
### 自定义 DHCP 选项

有些选项的内容，RFC2132 中没有统一规定，例如 Option43。下面将介绍设备上定义的几种选项格式。

- 厂商特定信息选项（Option43）

  Option43 称为厂商特定信息选项。Option43 的报文格式如图 6-3 所示。

**图 6-3** option43 格式




DHCP 服务器和 DHCP 客户端通过 Option43 交换厂商特定的信息。当 DHCP 服务器接收到请求 Option43 信息的 DHCP 请求报文（Option55 中带有 43 参数）后，将在回复报文中携带 Option43，为 DHCP 客户端分配厂商指定的信息。设备作为 DHCP 客户端时，可以通过 Option43 获取以下信息：

- ACS（Auto-Configuration Server，自动配置服务器）的参数，包括 URL 地址、用户名和密码。
- 服务提供商标识，CPE（Customer Premises Equipment，用户侧设备）从 DHCP 服务器获取该信息后，将该信息通告给 ACS，以便 ACS 选择服务提供商特有的配置和参数等。
- PXE（Preboot eXecution Environment，预启动执行环境）引导服务器地址，以便客户端从 PXE 引导服务器获取启动文件或其他控制信息。

为了提供可扩展性，通过 Option43 为客户端分配更多的信息，Option43 采用子选项的形式，通过不同的子选项为用户分配不同的网络配置参数，如图 6-3 所示。子选项中各字段的含义为：

- Sub-option type：子选项类型。目前，子选项类型值可以为 0x01 表示 ACS 参数子选项，0x02 表示服务提供商标识子选项，0x80 表示 PXE 引导服务器地址子选项。
- Sub-option length：子选项的长度。
- Sub-option value：子选项的取值。

- 中继代理信息选项（Option82）

  Option82 称为中继代理信息选项，该选项记录了 DHCP 客户端的位置信息。DHCP 中继或 DHCP Snooping 设备接收到 DHCP 客户端发送给 DHCP 服务器的请求报文后，在该报文中添加 Option82，并转发给 DHCP 服务器。

  管理员可以从 Option82 中获得 DHCP 客户端的位置信息，以便定位 DHCP 客户端，实现对客户端的安全和计费等控制。支持 Option82 的服务器还可以根据该选项的信息制定 IP 地址和其他参数的分配策略，提供更加灵活的地址分配方案。

  Option82 最多可以包含 255 个子选项。若定义了 Option82，则至少要定义一个子选项。目前设备只支持两个子选项：sub-option1（Circuit ID，电路 ID 子选项）和 sub-option2（Remote ID，远程 ID 子选项）。

  由于 Option82 的内容没有统一规定，不同厂商通常根据需要进行填充。

  AR200-S 提供了三种系统预定义的 Option82 格式，分别是 Default、Common、extend，还支持自定义格式 User-defined。

  - Default 格式是默认格式，这个是 Option82 的默认处理方式。
  - Common 格式是满足特定市场而对 Option82 做的配置，而且是全字符串的方式。
  - Extend 格式主要是兼容其他厂商路由器的 Option82 的格式，同时满足二进制的 Option82 表示方式。
  - User-defined：其它未特定描述的格式可以使用自定义的方式来处理。

## 6.4.3 DHCP Client

DHCP Client 即 DHCP 客户端，使用 DHCP 协议从 DHCP Server 获取 IP 地址和配置信息。

AR200-S 支持在接口上配置 DHCP Client 功能，这样接口可以作为 DHCP Client，使用 DHCP 协议从 DHCP 服务器动态获得 IP 地址等参数，方便用户配置，也便于集中管理。

说明

AR200-S 只支持在三层物理接口上配置 DHCP Client 功能，不支持二层接口和 VLANIF 接口。

## 6.4.4 DHCP Server

### 特点

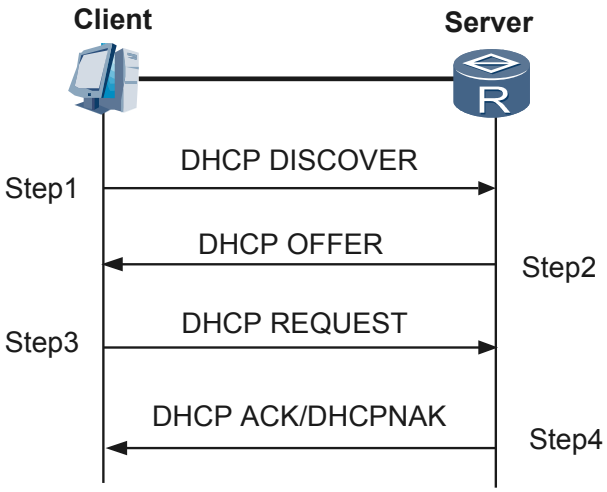
DHCP Server 即 DHCP 服务器，负责客户端 IP 地址的分配。客户端向服务器发送配置申请报文（包括 IP 地址、子网掩码、缺省网关等参数），服务器根据策略返回携带相应配置信息的报文给客户端。请求报文和回应报文都采用 UDP 进行封装。

### DHCP 客户端与服务器的交互模式

DHCP 客户端为了获取合法的动态 IP 地址，在不同阶段与服务器之间交互不同的信息，通常存在以下三种模式（根据 RFC2131 实现）：

- DHCP 客户端首次登录网络

**图 6-4 首次登录网络时 DHCP 客户端与 DHCP 服务器的交互过程**




如图 6-4 所示，DHCP 客户端首次登录网络时，主要通过四个阶段与 DHCP 服务器建立联系。

- 发现阶段，即 DHCP 客户端寻找 DHCP 服务器的阶段。客户端以广播方式发送 DHCP DISCOVER 报文，只有 DHCP 服务器才会进行响应。
- 提供阶段，即 DHCP 服务器提供 IP 地址的阶段。DHCP 服务器接收到客户端的 DHCP DISCOVER 报文后，从 IP 地址池中挑选一个尚未分配的 IP 地址分配给客户端，向该客户端发送包含出租 IP 地址和其它设置的 DHCP OFFER 报文。

- 选择阶段，即 DHCP 客户端选择 IP 地址的阶段。如果有多台 DHCP 服务器向该客户端发来 DHCP OFFER 报文，客户端只接收第一个收到的 DHCP OFFER 报文，然后以广播方式向各 DHCP 服务器回应 DHCP REQUEST 报文，该信息中包含向所选定的 DHCP 服务器请求 IP 地址的内容。
- 确认阶段，即 DHCP 服务器确认所提供 IP 地址的阶段。当 DHCP 服务器收到 DHCP 客户端回答的 DHCP REQUEST 报文后，便向客户端发送包含它所提供的 IP 地址和其它设置的 DHCP ACK 确认报文。DHCP 客户端收到该确认报文后，会以广播的方式发送免费 ARP 报文，探测是否有主机使用服务器分配的 IP 地址，如果在规定的时间内没有收到回应，客户端才使用此地址。否则，客户端会发送 DHCP DECLINE 报文给 DHCP 服务器，通知 DHCP 服务器该地址不可用，并重新申请 IP 地址。

除 DHCP 客户端选中的服务器外，其它 DHCP 服务器本次未分配出的 IP 地址仍可用于其他客户端的 IP 地址申请。

- DHCP 客户端再次登录网络

  当 DHCP 客户端再次登录网络时，主要通过以下几个步骤与 DHCP 服务器建立联系：

  - DHCP 客户端首次正确登录网络后，以后再登录网络时，只需要广播包含上次分配 IP 地址的 DHCP REQUEST 报文即可，不需要再次发送 DHCP DISCOVER 报文。
  - DHCP 服务器收到 DHCP REQUEST 报文后，如果客户端申请的地址没有被分配，则返回 DHCP ACK 确认报文，通知该 DHCP 客户端继续使用原来的 IP 地址。
  - 如果此 IP 地址无法再分配给该 DHCP 客户端使用（例如已分配给其它客户端），DHCP 服务器将返回 DHCP NAK 报文。客户端收到后，重新发送 DHCP DISCOVER 报文请求新的 IP 地址。

- DHCP 客户端延长 IP 地址的租用有效期

  DHCP 服务器分配给客户端的动态 IP 地址通常有一定的租借期限，期满后服务器会收回该 IP 地址。如果 DHCP 客户端希望继续使用该地址，需要更新 IP 租约（如延长 IP 地址租约）。

  DHCP 客户端延长 IP 地址的租用有效期的具体过程参见**地址池的租期**一节的描述。

## IP 地址的静态和动态分配

对于 IP 地址的占用时间，不同主机有不同的需求：对于服务器，可能需要长期使用确定的 IP 地址；对于某些主机，可能需要长期使用某个 IP 地址；而某些个人则可能只在需要时动态分配一个临时的 IP 地址就可以了。

针对这些不同的需求，DHCP 服务器提供三种 IP 地址分配策略：

- 手工分配地址：由管理员为少数特定主机（如 WWW 服务器等）配置固定的 IP 地址。
- 自动分配地址：为首次连接到网络的某些主机分配固定 IP 地址，该地址将长期由该主机使用。
- 动态分配地址：以“租借”的方式将某个地址分配给客户端主机，使用期限到期后，客户端需要重新申请地址。绝大多数客户端主机得到的是这种动态分配的地址。

DHCP 服务器按照如下次序为客户端选择 IP 地址：

1. DHCP 服务器的数据库中与客户端 MAC 地址静态绑定的 IP 地址；

2. 客户端以前曾经使用过的 IP 地址，即客户端发送的 DHCP DISCOVER 报文中请求 IP 地址选项（Requested IP Addr Option）的地址；
3. 在 DHCP 地址池中，顺序查找可供分配的 IP 地址，最先找到的 IP 地址；
4. 如果在 DHCP 地址池中未找到可供分配的 IP 地址，则依次查询超过租期、发生冲突的 IP 地址，如果找到可用的 IP 地址，则进行分配。

## 防止 IP 地址重复分配的方法

为防止 IP 地址重复分配导致地址冲突，DHCP 服务器为客户端分配地址前，需要先对该地址进行探测。

地址探测是通过 Ping 命令实现的，检测是否能在指定时间内得到 Ping 应答。如果没有得到应答，则继续发送 Ping 报文，直到发送 Ping 包的数量达到最大值。如果仍然超时，则可以认为这个 IP 地址的网段内没有设备使用该 IP 地址，从而确保客户端被分得的 IP 地址是唯一的（根据 RFC2132 实现）。

缺省情况下，DHCP 服务器发送 ping 报文数量为 0，即不进行 ping 操作，等待 Ping 响应的最长时间为 500 毫秒。

## IP 地址预留

DHCP 支持预留 IP 地址给客户端，预留的 IP 地址可以是地址池中的地址，也可以不是。如果是，则将其不列入到地址池可分配的 IP 地址中，这些预留出去的地址一般用于某个 DNS 服务器的 IP 地址。

## 地址池的租期

对于不同的地址池，DHCP 服务器可以指定不同的地址租用期限，但同一 DHCP 地址池中的地址都具有相同的期限。

DHCP 服务器分配给客户端的动态 IP 地址通常有一定的租借期限，期满后服务器会收回该 IP 地址。如果 DHCP 客户端希望继续使用该地址，需要更新 IP 租约（如延长 IP 地址租约）。

当 DHCP 客户端获得 IP 地址时，会进入到绑定状态，客户端会设置 3 个定时器，分别用来控制租期更新、重绑定和判断是否已经到达租期。DHCP 为客户分配 IP 地址时，可以为定时器指定确定的值。若服务器没有设置定时器的值，客户端就使用缺省值。定时器的缺省值如**表 6-1** 所示。

**表 6-1 定时器的缺省值**

| 定时器 | 默认值 |
| --- | --- |
| 租期更新 | 总租期的 50% |
| 重绑定 | 总租期的 87.5% |
| 到达租期 | 总租期 |

当“租期更新定时器”到期时，DHCP 客户端必须进行 IP 地址的更新。DHCP 客户端会自动地向曾经为自己分配过 IP 地址的 DHCP 服务器发送 DHCP REQUEST 报文，此时客户端进入到更新状态，如果此 IP 地址有效，则 DHCP 服务器回应 DHCP ACK 报文，

通知 DHCP 客户端已经获得新 IP 租约，客户端会重新进入到绑定状态。若客户端收到 DHCP 服务器返回的 DHCP NAK 报文，则进入到初始化状态。

当客户端发送延长租期的 DHCP REQUEST 报文后，保持在更新状态等待响应。若直到“重绑定定时器”到期，客户端还没有收到服务器的响应，客户端会假定原来的 DHCP 服务器不可用，并开始发送广播的 DHCP REQUEST 报文。

网络上的任何 DHCP 服务器均可以响应此客户端的请求，并向此客户端发送 DHCP ACK 报文或者 DHCP NAK 报文。

如果客户端收到一个 DHCP ACK 报文，那么就返回到绑定状态，且重新设置“租期更新和重绑定定时器”，如果客户端收到的都是 DHCP NAK 报文，那么就返回到初始化状态。此时客户端必须立即停止使用此 IP 地址，并且返回到初始化状态，重新申请新的 IP 地址。

若客户端在“到达租期定时器”到期前都没有收到响应，客户端必须立即停止使用此 IP 地址，并且返回到初始化状态，重新发送 DHCP DISCOVER 报文请求新的 IP 地址（根据 RFC2131 实现）。

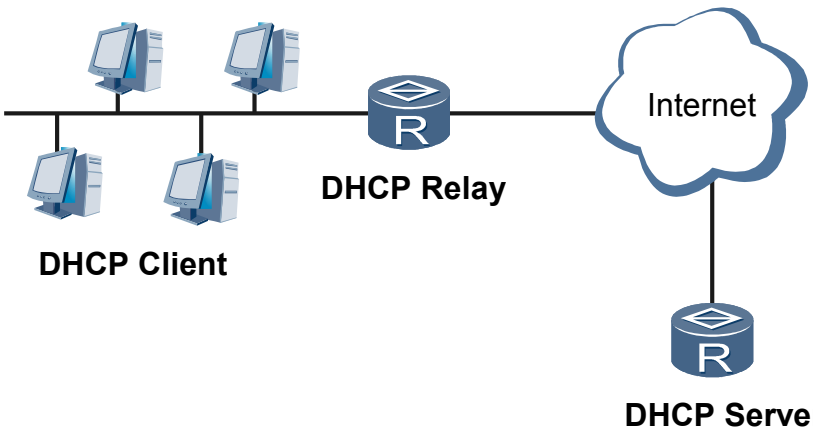
## 6.4.5 DHCP Relay

### 应用环境

由于在 IP 地址动态获取过程中采用广播方式发送请求报文，因此 DHCP 只适用于 DHCP 客户端和服务器处于同一个子网内的情况。为进行动态主机配置，需要在所有网段上都设置一个 DHCP 服务器，这显然是很不经济的。

DHCP 中继功能的引入解决了这一难题：客户端可以通过 DHCP 中继与其他网段的 DHCP 服务器通信，最终获取到 IP 地址。这样，多个网络上的 DHCP 客户端可以使用同一个 DHCP 服务器，既节省了成本，又便于进行集中管理。DHCP 中继的应用环境如图 6-5 所示。

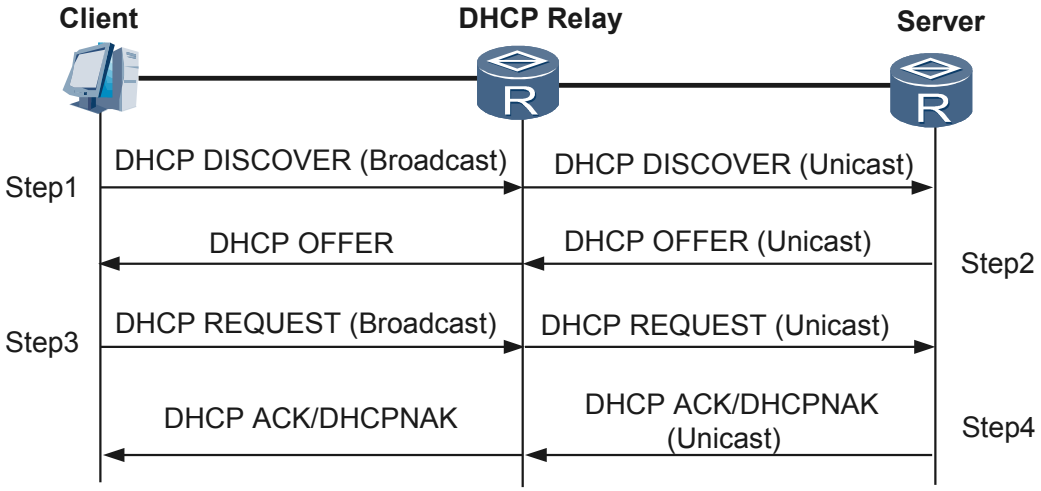
**图 6-5** DHCP 中继的应用环境




### 基本原理

DHCP 客户端与 DHCP 服务器通过 DHCP 中继的交互过程如图 6-6 所示。

**图 6-6** DHCP 客户端与 DHCP 服务器通过 DHCP 中继的交互过程




DHCP 中继的工作过程为：

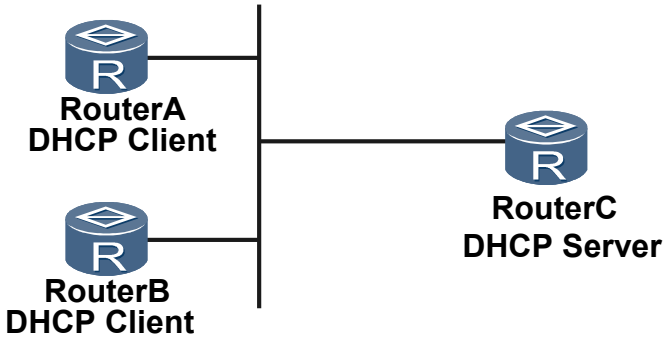
- 具有 DHCP 中继功能的网络设备收到 DHCP 客户端以广播方式发送的 DHCP-DISCOVER 或 DHCP-REQUEST 报文后，将报文中的 giaddr 字段填充为 DHCP 中继的 IP 地址，并根据配置将报文单播转发给指定的 DHCP 服务器。
- DHCP 服务器根据 giaddr 字段为客户端分配 IP 地址等参数，并通过 DHCP 中继将配置信息转发给客户端，完成对客户端的动态配置。

# 6.5 应用

## 6.5.1 DHCP Client 的典型组网应用

DHCP Client 的典型组网如图 6-7 所示。

**图 6-7** AR200-S 接口作为 DHCP Client 的典型组网




AR200-S 支持在接口上配置 DHCP Client 功能，如图 6-7 所示。这样接口可以作为 DHCP Client，使用 DHCP 协议从 DHCP 服务器动态获得 IP 地址等参数，方便用户配置，也便于集中管理。

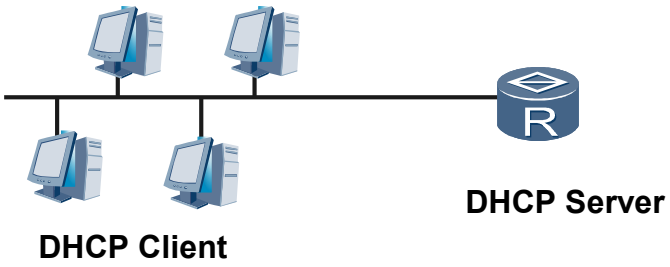
说明

AR200-S 只支持在三层物理接口上配置 DHCP Client 功能，不支持二层接口和 VLANIF 接口。

## 6.5.2 DHCP Server 的典型组网应用

DHCP 服务器典型应用如图 6-8 所示。

**图 6-8** DHCP Server 的应用环境



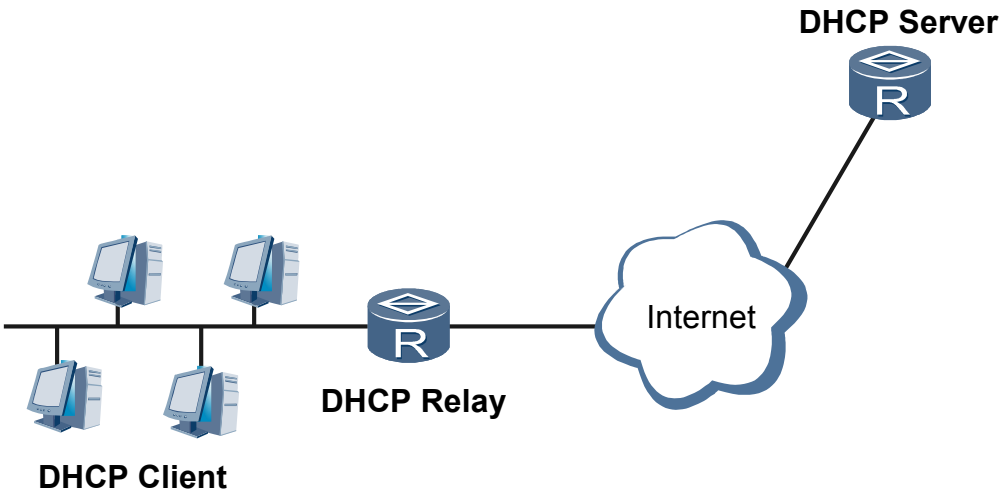

在以下场合通常利用 DHCP 服务器来完成 IP 地址分配。

- 网络规模较大，手工配置需要很大的工作量，并难以对整个网络进行集中管理。
- 网络中主机数目大于该网络支持的 IP 地址数量，无法给每个主机分配一个固定的 IP 地址。大量用户必须通过 DHCP 服务动态获得自己的 IP 地址，而且，对并发用户的数目也有限制。
- 网络中需要固定 IP 地址的主机比较少，大部分主机可以不使用固定的 IP 地址。

## 6.5.3 DHCP Relay 的典型组网应用

DHCP 中继的典型应用如图 6-9 所示。

**图 6-9** DHCP 中继的应用环境



DHCP 中继的应用环境：早期的 DHCP 协议只适用于 DHCP 客户端和服务器处于同一个网段内的情况，不能跨网段。因此，为进行动态主机配置，需要在每个网段置一个 DHCP 服务器，这显然是很不经济的。

DHCP 中继（DHCP Relay）功能的引入解决了这一难题：客户端可以通过 DHCP 中继与其他网段的 DHCP 服务器通信，最终取得合法的 IP 地址。这样，多个网段的 DHCP 客户端可以使用同一个 DHCP 服务器，既节省了成本，又便于进行集中管理。

一般来说，DHCP 中继既可以是主机，也可以是三层交换机，或者是路由器，只要在设备上启动 DHCP 中继代理的服务程序即可。

# 6.6 术语与缩略语

## 缩略语

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| DHCP | Dynamic Host Configure Protocol | 动态主机配置协议 |

# 7 NAT

## 关于本章

# 7.1 介绍

NAT 是将 IP 数据报报头中的 IP 地址转换为另一个 IP 地址的过程，主要用于实现内部网络（私有 IP 地址）访问外部网络（公有 IP 地址）的功能。

在实际应用中，内部网络一般使用私有地址。RFC（Request For Comments）1918 为私有地址留出了三个 IP 地址块。具体如下：

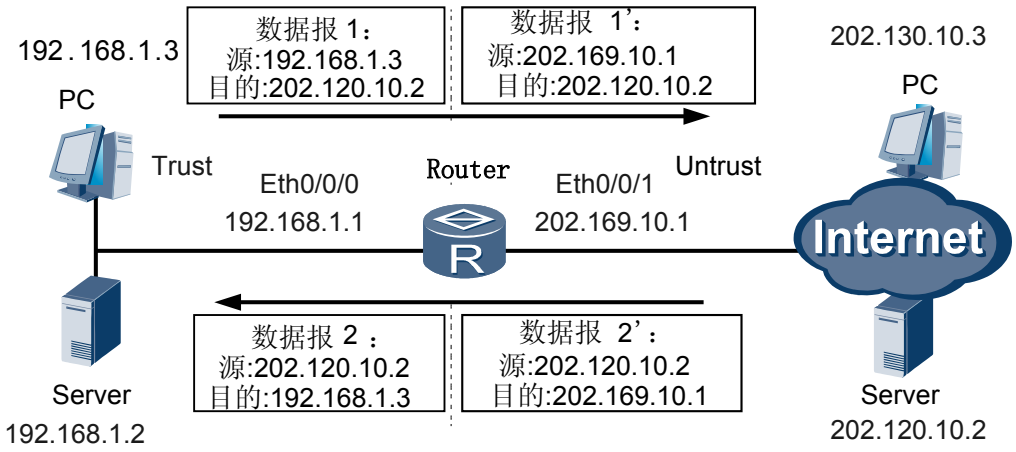
- A 类 10.0.0.0 ～ 10.255.255.255（10.0.0.0/8）
- B 类 172.16.0.0 ～ 172.31.255.255（172.16.0.0/12）
- C 类 192.168.0.0 ～ 192.168.255.255（192.168.0.0/16）

上述三个范围内的地址不会在 Internet 上被分配，因而可以不必向 ISP（Internet Service Provider）或注册中心申请而在公司或企业内部自由使用。

NAT 主要用于实现私有网络访问外部网络的功能。通过应用 NAT，能够使多数的私有 IP 地址转换为少数的公有 IP 地址，减缓可用 IP 地址空间枯竭的速度。

图 7-1 描述了一个基本的 NAT 应用。

**图 7-1 地址转换的基本过程**




NAT 服务器处于私有网络和公有网络的连接处，内部 PC 与外部服务器的交互报文全部通过该 NAT 服务器。地址转换的过程如下。

1. 内部 PC（192.168.1.3）发往外部服务器（202.120.10.2）的数据报 1 到达 NAT 服务器后，NAT 服务器查看报头内容，发现该数据报为发往外部网络的报文。
2. NAT 服务器将数据报 1 的源地址字段的私有地址 192.168.1.3 换成一个可在 Internet 上选路的公有地址 202.169.10.1，发送到外部服务器，同时在网络地址转换表中记录这一地址转换映射。
3. 外部服务器收到数据报 1’后，向内部 PC 发送应答报文，即数据报 2’，初始目的地址为 202.169.10.1。
4. 数据报 2’到达 NAT 服务器后，NAT 服务器查看报头内容，查找当前网络地址转换表的记录，用私有地址 192.168.1.3 替换目的地址，发送给内部 PC。

上述的 NAT 过程对 PC 和外部服务器来说是透明的。内部 PC 认为与外部服务器的交互报文没有经过 NAT 服务器的干涉；外部服务器认为内部 PC 的 IP 地址就是 202.169.10.1，并不知道存在 192.168.1.3 这个地址。

# 7.2 参考标准和协议

本特性的参考资料清单如下：

| 文档 | 描述 |
| --- | --- |
| RFC 1631 | The IP Network Address Translator (NAT) |
| RFC 2663 | IP Network Address Translator (NAT) Terminology and Considerations |
| RFC 2709 | Security Model with Tunnel-mode IPsec for NAT Domains |
| RFC 2766 | Network Address Translation - Protocol Translation (NAT-PT) |
| RFC 2993 | Architectural Implications of NAT |
| RFC 3022 | Traditional IP Network Address Translator (Traditional NAT) |
| RFC 3235 | Network Address Translator (NAT)-Friendly Application Design Guidelines |
| RFC 3519 | Mobile IP Traversal of Network Address Translation (NAT) Devices |
| RFC 3715 | IPsec-Network Address Translation (NAT) Compatibility Requirements |
| RFC 3947 | Negotiation of NAT-Traversal in the IKE |
| RFC 4008 | Definitions of Managed Objects for Network Address Translators (NAT) |
| RFC 4787 | Network Address Translation (NAT) Behavioral Requirements for Unicast UDP |

# 7.3 可获得性

## 涉及网元

无需其他网元的配合。

## License 支持

无需获得 License 许可，即可获得该特性的服务。

### 版本支持

| 产品 | 支持版本 |
| --- | --- |
| AR200-S | V200R002C00 |

### 特性依赖

不依赖其他特性。

### 硬件要求

对硬件无特殊要求。

## 7.4 原理描述

### 7.4.1 NAT 的转换机制

NAT 地址转换的机制分为如下两个部分：

- 内部网络主机的 IP 地址和端口转换为 AR200-S 的外部网络地址和端口。
- 外部网络地址和端口转换为 AR200-S 的内部网络主机的 IP 地址和端口。

也就是私有地址及其端口与公有地址及其端口之间的转换。

当数据流从一个安全区域流向另一个安全区域时，AR200-S 将检测该数据连接是否需要进行 NAT 转换。如果需要，则按照如下原则进行 NAT 转换：

- 在 IP 转发的出口，AR200-S 将报文的源地址（私有地址）转换为公网地址，并向外部网络发送。
- 在 IP 转发的入口，AR200-S 将报文的目的地址（公网地址）转换为私网地址，并向内部网络发送。

### 7.4.2 NAT 地址转换

#### 多对多地址转换及控制

NAT 允许 NAT 服务器拥有多个公有 IP 地址，实现了并发性。当第一个内部主机访问外部网络时，AR200-S 选择公有地址 IP1 作为其公网 IP 地址；当另一内部主机访问外部网络时，AR200-S 选择公有地址 IP2 作为其公网 IP 地址。以此类推，从而满足多台内部主机访问外部网络的请求。这种 NAT 转换称为“多对多地址转换”。

说明

NAT 服务器拥有的公有 IP 地址数目要远少于内部网络的主机数目，因为所有内部主机并不会同时访问外部网络。公有 IP 地址数目的确定，应根据网络高峰期可能访问外部网络的内部主机数目的统计值来确定。

在实际应用中，用户可能希望某些内部的主机具有访问 Internet 的权利，而某些主机没有。即当 NAT 进程查看数据报报头内容时，如果发现源 IP 地址是为那些不允许访问外

部网络的内部主机所拥有的，将不进行 NAT 转换。这就是一个对地址转换进行控制的问题。

AR200-S 是通过定义地址池来实现多对多地址转换，同时利用访问控制列表来对地址转换进行控制的。具体解释如下：

- 地址池是用于地址转换的一些公有 IP 地址的集合。用户应根据自己拥有的合法 IP 地址数目、内部网络主机数目以及实际应用情况，配置恰当的地址池。地址转换的过程中，将会从地址池中挑选一个地址做为转换后的源地址。
- 利用访问控制列表限制地址转换并关联地址池。当内部网络有数据包要发往外部网络时，AR200-S 首先根据访问控制列表判定是否是允许的数据包，然后根据关联关系，找到对应的地址池，把源地址转换成这个地址池中的某一个地址。这可以有效地控制地址转换的使用范围，使特定主机能够有权访问 Internet。

AR200-S 在处理报文时，如果发现该报文是需要从内网发往外网的数据，并且允许进行地址转换，根据“转换关联”可以利用地址池将内部网络主机的 IP 地址和端口替换为 AR200-S 的外部网络地址和端口，并且将转换关系记录到会话表中；对于外网需要发往内网的数据，通过查找已经建立的会话表，把 AR200-S 的外部网络地址和端口转换为内部网络主机的 IP 地址和端口。

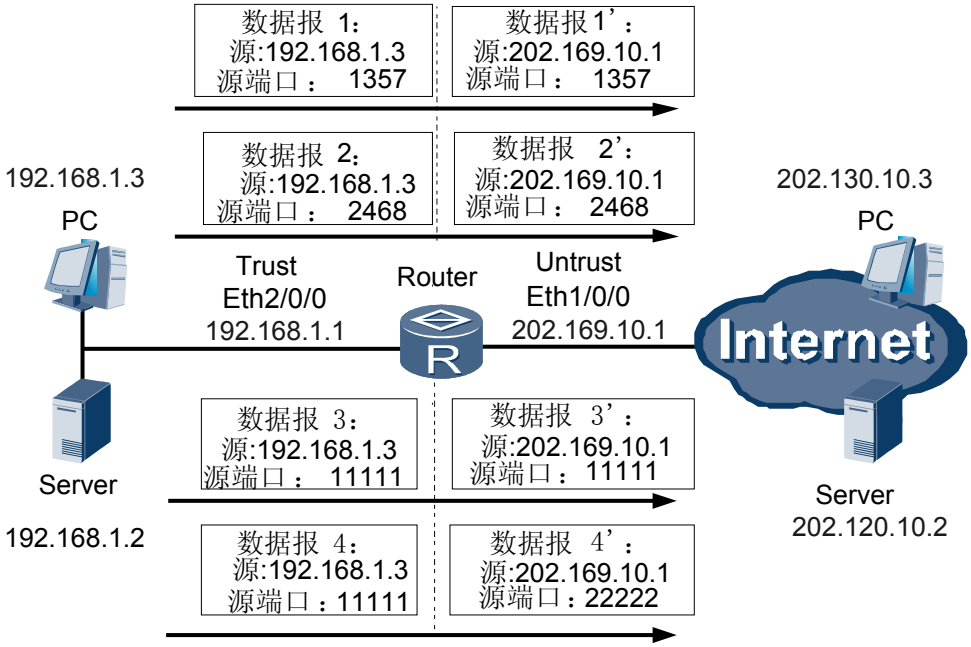
## NAPT

除多对多的 NAT 转换方式外，NAPT（Network Address Port Translation）也能实现并发的地址转换。它允许多个内部地址映射到同一个公有地址上，非正式的也可称之为“多对一地址转换”或地址复用。

NAPT 映射 IP 地址和端口号，来自不同内部地址的数据报可以映射到同一公有地址的不同端口号上，因而仍然能够共享同一公有地址。

图 7-2 描述了 NAPT 的基本原理。

**图 7-2** NAPT 示意图



如图 7-2 所示，四个带有内部地址的数据报到达 NAT 服务器。其中：

- 数据报 1 和数据报 2 来自同一个内部地址但有不同的源端口号。
- 数据报 3 和数据报 4 来自不同的内部地址但具有相同的源端口号。

通过 NAT 映射，四个数据报都被转换到同一个外部地址的不同源端口号上，因而仍保留了报文之间的区别。

当回应报文到达时，NAT 进程仍能够根据回应报文的目的地址和端口号区别该报文，并将返回报文 NAT 转换后，转发给相应内部主机。

配置 NAPT 功能后，NAT 转换时，AR200-S 首先将复用地址池中所选择的地址，端口达到能力极限后，再选择另一个地址完成转换。对比单一的多对多地址转换，这可大大减少地址池中公有地址的数目。

### 静态 NAT/NAPT

静态 NAT，是指在进行 NAT 时，内部网络主机的 IP 同公网 IP 是一对一静态绑定的，静态 NAT 中的公网 IP 只会给唯一且固定的内网主机 IP 转换使用。

静态 NAPT，是指“内部网络主机的 IP+协议号+端口号”同“公网 IP+协议号+端口号”是一对一静态绑定的，静态 NAPT 中的公网 IP 可以为多个 NAPT 条目使用。

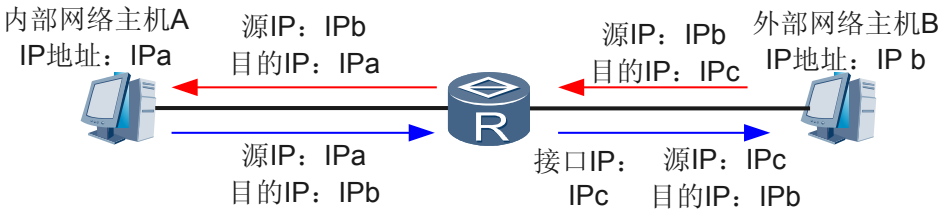
静态 NAT/NAPT，支持内网主机发起对外网主机的访问，以及外网主机发起对内网主机的访问。

静态 NAT/NAPT，还支持将指定范围内的内部主机 IP 转换为指定的公网网段 IP，转换过程中只对网段地址进行转换，保持主机地址不变。当内部主机访问外部网络时，如果主机地址在指定的内部主机地址范围内，会被转换为对应的公网地址；同样，当通过公网网段地址对内部主机进行访问时，可以直接访问到内部主机（该公网地址在转换后在指定的内部主机地址范围内）。

## 7.4.3 Easy IP

AR200-S 在做传统的源 NAT 时，需要用户配置额外的 NAT 地址池，地址池中的地址都是公网地址，在内部网络规模小、公网地址资源有限的情况下，如果直接使用 AR200-S 对外接口的公网 IP 地址作为 NAT 的源地址，就可以节省公网地址资源，地址转换流程如图 7-3 所示：

**图 7-3** Easy IP 示意图




同样，它也利用访问控制列表控制哪些内部地址可以进行地址转换。

Easy IP 方式特别适合小型局域网访问 Internet 的情况。这里的小型局域网主要指中小型网吧、小型办公室等环境，一般具有内部主机较少的特点。对于这种情况，可以使用 Easy IP 方式使局域网用户都通过接口的公网 IP 地址接入 Internet。

## 7.4.4 NAT server

NAT 隐藏了内部网络的结构，具有“屏蔽”内部主机的作用，但是在实际应用中，可能需要提供给外部一个访问内部主机的机会，如提供给外部一台 WWW 服务器，或是一台 FTP 服务器。使用 NAT server 功能可以灵活地添加内部服务器。AR200-S 提供两种方式为内部服务器指定外部地址，例如：

- 可以使用 202.169.10.10 作为 Web 服务器的外部地址。
- 可以使用 202.110.10.12:8080 作为 Web 服务器的外部地址。

AR200-S 的 NAT server 功能能够为外部网络用户提供访问的内部服务器。外部用户访问内部服务器时，有如下两部分操作：

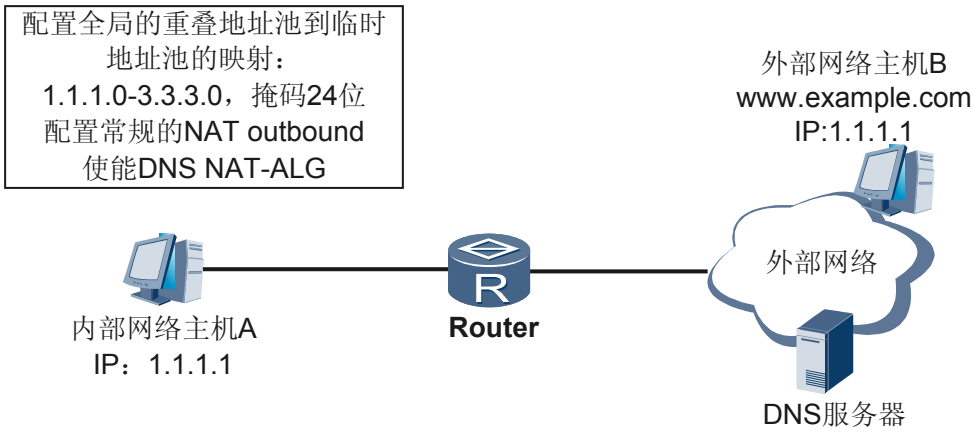
- AR200-S 将外部用户的请求报文的目的地址转换成内部服务器的私有地址。
- AR200-S 将内部服务器的回应报文的源地址（私网地址）转换成公网地址。

AR200-S 支持为外部用户提供多台同样的服务器，例如，提供多台 Web 服务器。

## 7.4.5 两次 NAT

两次 NAT 即 Twice NAT，指源 IP 和目的 IP 同时转换，该技术应用于内部网络主机地址与外部网络上主机地址重叠的情况。

**图 7-4** 两次 NAT 示意图



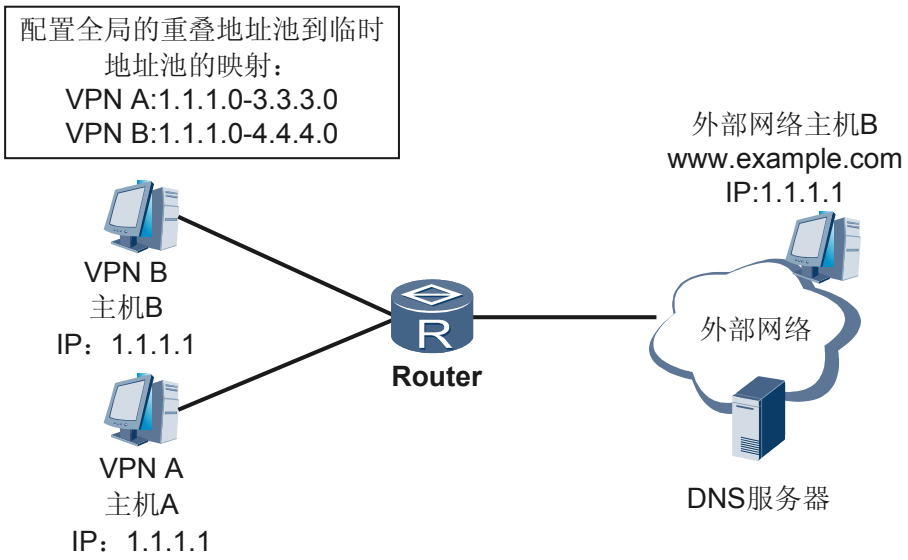

如图 7-4 示，用户在 AR200-S 上配置重叠地址池到临时地址的映射关系,同时配置常规的源 NAT，地址转换过程如下:

1. 内网主机 A 要访问地址重叠的外部网络主机 B，主机 A 向位于外部网络的 DNS 服务器发送访问外网主机 B 的 DNS 请求，DNS 服务器应答主机 B 的 IP 地址为 1.1.1.1，DNS 应答报文在经过 AR200-S 时，进行 DNS ALG，AR200-S 将 DNS 应答报文中的重叠地址 1.1.1.1 转换为唯一的临时地址 3.3.3.1，然后再转发给主机 A。
2. 主机 A 访问主机 B，目的 IP 为临时地址 3.3.3.1，报文在经过 AR200-S 时，AR200-S 检查到目的 IP 是临时地址，进行目的 NAT，将报文的目的 IP 转换为主机 B 的真实地址 1.1.1.1，同时进行正常的 NAT outbound，将报文的源 IP 转换为源 NAT 地址池地址；AR200-S 将报文转发到主机 B。

3. 主机 B 回应主机 A，目的 IP 为主机 A 的 NAT outbound 地址池地址，源 IP 为主机 B 的地址 1.1.1.1，报文在经过 AR200-S 时，AR200-S 检查到源 IP 是重叠地址，进行源 NAT，将报文的源 IP 转换为对应的临时地址 3.3.3.1，同时进行正常的目的 NAT，将报文的目的 IP 从源 NAT 地址池地址转换为主机 A 的内网地址 1.1.1.1；AR200-S 将报文转发到主机 A。

考虑到内网有多个 VPN 的场景，且内网多个 VPN 的地址一样的情况下，在 AR200-S DNS ALG 时，增加内网 VPN 信息作为重叠地址池到临时地址的映射关系匹配条件之一，如图 7-5 所示：

**图 7-5** 内网多 VPN 情况下的两次 NAT 示意图

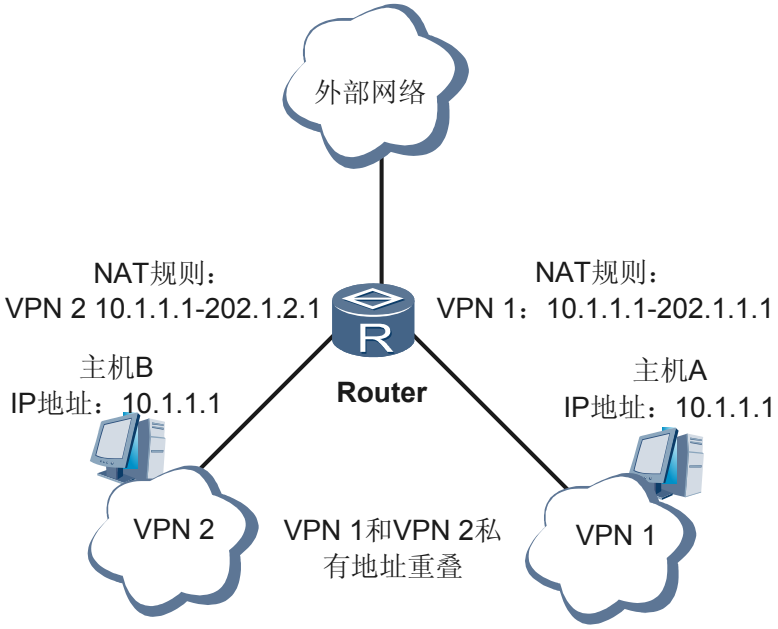



## 7.4.6 VPN 关联的源 NAT

AR200-S 的 NAT 不仅可以使内部网络的用户访问外部网络，还允许分属于不同 VPN（Virtual Private Network）的用户通过同一个出口访问外部网络，能够解决内部网络中 IP 地址重叠的 VPN 同时访问外网主机的问题。

场景如图 7-6 所示：

**图 7-6** VPN 关联的源 NAT




访问流程如下所述：

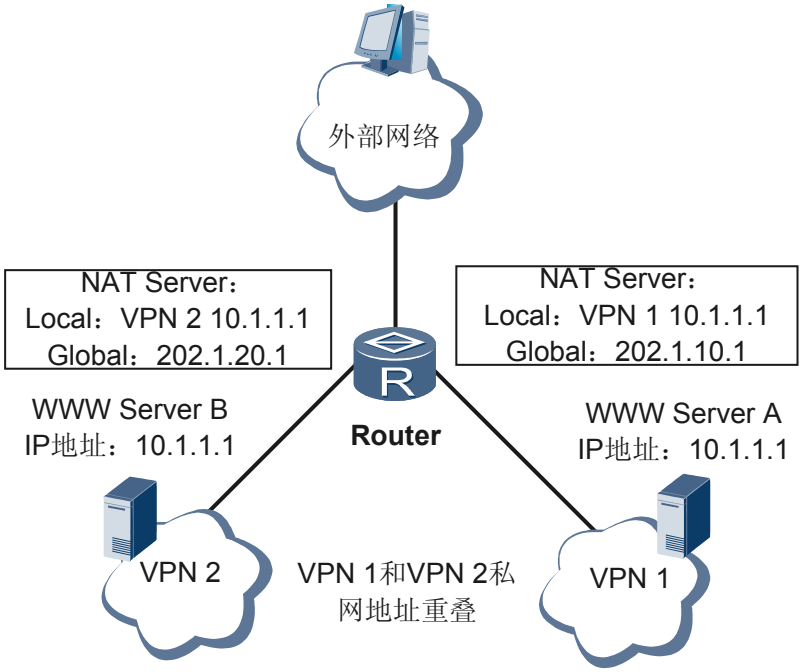
1. VPN 1 内的主机 A 和 VPN 2 内的主机 B 地址重叠，都为私网地址 10.1.1.1，都要同时访问外部网络的一个服务器。
2. AR200-S 在做源 NAT 时，将内部 VPN 作为一个 NAT 的匹配条件，将主机 A 发出报文的源 IP 转换为 202.1.1.1，将主机 B 发出报文的源 IP 转换为 202.1.2.1，同时在建立的 NAT 会话表中，记录用户的 VPN 信息。
3. 当外部网络服务器回应内部网络主机 A 和 B 的报文经过 AR200-S 时，根据已建立的 NAT 会话表，NAT 模块将发往主机 A 报文的目的 IP 从 202.1.1.1 转换为 10.1.1.1，再发往 VPN 1 的目的主机；将发往主机 B 报文的目的 IP 从 202.1.2.1 转换为 10.1.1.1，再发往 VPN 2 的目的主机。

## 7.4.7 VPN 关联的 NAT Server

AR200-S 的 NAT 模块支持 VPN 关联的 NAT Server，提供给外部网络访问 VPN 内主机的机会，能够支持内网多个 VPN 地址重叠的场景。

场景如图 7-7 所示：

**图 7-7** VPN 关联的 NAT Server




上图中，VPN 1 内 WWW 服务器 A 和 VPN2 内的 WWW 服务器 B 的地址都是 10.1.1.1；使用 202.1.10.1 做为 VPN 1 内的服务器 A 的外部地址，使用 202.1.20.1 做为 VPN 2 内的服务器 B 的外部地址。这样，外部网络的用户使用 202.1.10.1 就可以访问到 VPN 1 提供的 WWW 服务，使用 202.1.20.1 就可以访问 VPN 2 提供的 WWW 服务。

访问流程如下所述：

1. 外部网络的主机访问 VPN 1 内的服务器 A，报文目的 IP 是 202.1.10.1；访问 VPN 2 内的服务器 B，报文目的 IP 是 202.1.20.1。
2. AR200-S 在做 NAT server 时，根据报文的目的 IP 及 VPN 信息进行判断，将目的 IP 是 202.1.10.1 的报文的目的 IP 转换为 10.1.1.1，然后发往 VPN 1 的目的 WWW Server A；将目的 IP 是 202.1.20.1 的报文的目的 IP 转换为 10.1.1.1，然后发往 VPN 2 的目的 WWW Server B；同时在新建的 NAT 会话表中，记录下关联的 VPN 信息。
3. 当内部 WWW 服务器 A 和 B 回应外部网络主机的报文经过 AR200-S 时，根据已建立的 NAT 会话表，NAT 模块将从服务器 A 发出的报文的源 IP 从 10.1.1.1 转换为 202.1.10.1，再发往外部网络；将从服务器 B 发出的报文的源 IP 从 10.1.1.1 转换为 202.1.20.1，再发往外部网络。

## 7.4.8 ALG-应用层网关

NAT 和 NAPT 只能对 IP 报文的头部地址和 TCP/UDP 头部的端口信息进行转换。对于一些特殊协议，例如 ICMP、FTP 等，它们报文的数据部分可能包含 IP 地址或端口信息，这些内容不能被 NAT 有效的转换，就可能导致问题。

例如，一个使用内部 IP 地址的 FTP 服务器可能在和外部网络主机建立会话的过程中需要将自己的 IP 地址发送给对方。而这个地址信息是放到 IP 报文的数据部分，NAT 无法对它进行转换。当外部网络主机接收了这个私有地址并使用它，这时 FTP 服务器将表现为不可达。

解决这些特殊协议的 NAT 转换问题的方法就是在 NAT 实现中使用 ALG（Application Level Gateway）功能。ALG 是特定的应用协议的转换代理，它和 NAT 交互以建立状态，使用 NAT 的状态信息来改变封装在 IP 报文数据部分中的特定数据，并完成其他必需的工作以使应用协议可以跨越不同范围运行。

例如，考虑一个“目的站点不可达”的 ICMP 报文，该报文数据部分包含了造成错误的数据报 A 的首部（注意，NAT 发送 A 之前进行了地址转换，所以源地址不是内部主机的真实地址）。如果开启了 ICMP ALG 功能，在 NAT 转发 ICMP 报文之前，它将与 NAT 交互，打开 ICMP 报文并转换其数据部分的报文 A 首部的地址，使这些地址表现为内部主机的确切地址形式，并完成其他一些必需工作后，由 NAT 将这个 ICMP 报文转发出去。

AR200-S 提供了完善的地址转换应用级网关机制，使其在流程上可以支持各种特殊的应用协议，而不需要对 NAT 平台进行任何的修改，具有良好的可扩充性。目前它所实现的常用应用协议的 ALG 功能包括：

- DNS
- FTP
- ICMP
- SIP
- RTSP

ALG 功能对常用应用协议报文做 NAT 变换时，其报文中有部分字段会做 NAT 变换。

- DNS ALG 功能对应用协议报文做 NAT 变换时，其报文中做 NAT 变换的字段为：A 响应报文中的 IP 和 Port。
- FTP ALG 功能对应用协议报文做 NAT 变换时，其报文中做 NAT 变换的字段为：
  - Port 请求报文中载荷里的 IP 和 Port。
  - Passive 响应报文中载荷里的 IP 和 Port。
- ICMP ALG 功能对应用协议报文做 NAT 变换时，其报文中做 NAT 变换的字段为：ICMP 报文载荷部分的 IP 和 port。
- SIP ALG 功能对应用协议报文做 NAT 变换时，其报文中做 NAT 变换的字段为：
  - Request line
  - From
  - To
  - Contact
  - Via
  - O
  - Message body 的 C 字段地址和 M 字段的端口
- RTSP ALG 功能对应用协议报文做 NAT 变换时，其报文中做 NAT 变换的字段为：setup/reply OK 报文中的端口字段。

## 7.4.9 NAT 映射

在 Internet 中使用 NAT 映射功能，所有不同的信息流看起来好像来源于同一个 IP 地址。因为 NAT 映射使得一组主机可以共享唯一的外部地址，当位于内部网络中的主机通过 NAT 设备向外部主机发起会话请求时，NAT 设备就会查询 NAT 表，看是否有相关会话记录，如果有相关记录，就会将内部 IP 地址及端口同时进行转换，再转发出

去；如果没有相关记录，进行 IP 地址和端口转换的同时，还会在 NAT 表增加一条该会话的记录。NAT 映射是 NAT 设备对内网发到外网的流量进行映射。包括以下三种类型：

- 外部地址无关的映射：对相同的内部 IP 和端口重用相同的地址端口映射。
- 外部地址相关的映射：对相同的内部 IP 地址和端口访问相同的外部 IP 地址时重用相同的端口映射。
- 外部地址和端口相关的映射：对相同的内部 IP 地址和端口号访问相同的外部 IP 地址和端口号重用相同的端口映射（如果此映射条目还处在活动状态）。

AR200-S 支持外部地址无关、外部地址和端口相关的映射。

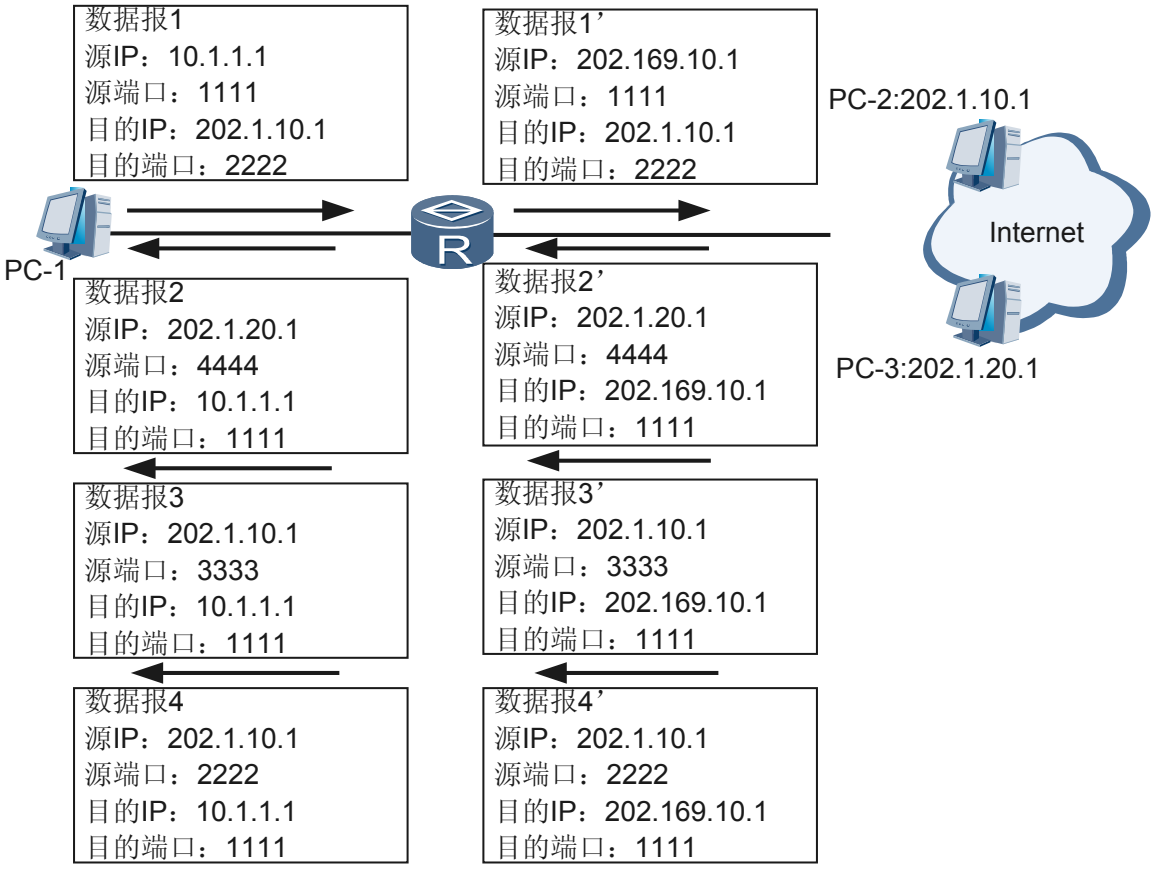
## 7.4.10 NAT 过滤

NAT 过滤是指 NAT 设备对外网发到内网的流量进行过滤。包括三种类型：

- 与外部地址无关的 NAT 过滤行为
- 与外部地址相关的 NAT 过滤行为
- 与外部地址和端口都相关的 NAT 过滤行为

场景应用如图 7-8 所示。

**图 7-8** NAT 过滤应用



上图中，私网用户 PC－1 通过 NAT 设备与外网用户 PC－2、PC－3 进行通信。数据报 1 代表私网主机 PC-1 访问公网 PC-2，PC-1 使用的端口号为 1111，访问 PC-2 的端口 2222；经过 NAT 设备时，源 IP 转换为 202.169.10.1。

当私网主机向某公网主机发起访问后，公网主机发向私网主机的流量经过 NAT 设备时需要进行过滤。数据报 2、数据报 3 和数据报 4 代表三种场景，分别对应上述三种 NAT 过滤类型。

- 数据报 2 代表公网主机 PC-3（与报文 1 的目的地址不同）访问私网主机 PC-1，目的端口号为 1111，只有配置了外部地址无关的 NAT 过滤行为，才允许此报文通过，否则被 NAT 设备过滤掉。
- 数据报 3 代表公网服务器 PC-2（与报文 1 的目的地址相同）访问私网主机 PC-1，目的端口号为 1111，源端口号为 3333（与报文 1 的目的端口不同），只有配置了外部地址相关的 NAT 过滤行为或者配置了外部地址无关的 NAT 过滤行为，才允许此报文通过，否则被 NAT 设备过滤掉。
- 数据报 4 代表公网服务器 PC-2（与报文 1 的目的地址相同）访问私网主机 PC-1，目的端口号为 1111，源端口号为 2222（与报文 1 的目的端口相同），这属于外部地址和端口都相关的 NAT 过滤行为，是缺省的过滤行为，不配置或者配置任何类型的 NAT 过滤行为，都允许此报文通过，不会被过滤掉。

AR200-S 支持如上三种类型的 NAT 过滤行为。

# 7.5 术语与缩略语

## 术语

| 术语 | 解释 |
| --- | --- |
| **N** | |
| **NAT** | 又称为地址代理，将私有地址转换为全球唯一公网地址，用于实现私有网络和公有网络之间的互访功能。 |
| | |
| **S** | |
| **私有地址** | 内部网络或主机的 IP 地址，不在 Internet 上被分配，可在一个单位或公司内部使用。 |

## 缩略语

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| **A** | | |
| **ACL** | Access Control List | 访问控制列表 |
| **ALG** | Application Layer Gateway | 应用层网关 |
| | | |

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| **F** | | |
| **FTP** | File Transfer Protocol | 文件传输协议 |
| | | |
| **H** | | |
| **HTTP** | Hyper Text Transport Protocol | 超文本传送协议 |
| | | |
| **I** | | |
| **ICMP** | Internet Control Message Protocol | 互联网控制报文协议 |
| **IGMP** | Internet Group Management Protocol | 互联网组管理协议 |
| **IP** | Internet Protocol | 互联网协议 |
| | | |
| **M** | | |
| **MAC** | Media Access Control | 媒体访问控制 |
| | | |
| **N** | | |
| **NAPT** | Network Address Port Translation | 网络地址端口转换 |
| **NAT** | Network Address Translation | 网络地址转换 |
| | | |
| **P** | | |
| **PAT** | Port Address Translation | 端口地址转换 |
| **PC** | Personal Computer | 个人计算机 |
| | | |
| **T** | | |
| **TCP** | Transmission Control Protocol | 传输控制协议 |
| **TTL** | Time to Live | 生存时间 |
| | | |
| **U** | | |
| **UDP** | User Datagram Protocol | 用户数据包协议 |
| | | |
| **V** | | |

| 缩略语 | 英文全称 | 中文全称 |
| --- | --- | --- |
| **VPN** | Virtual Private Network | 虚拟私有网 |
| | | |
| **W** | | |
| **WWW** | World Wide Web | 万维网 |