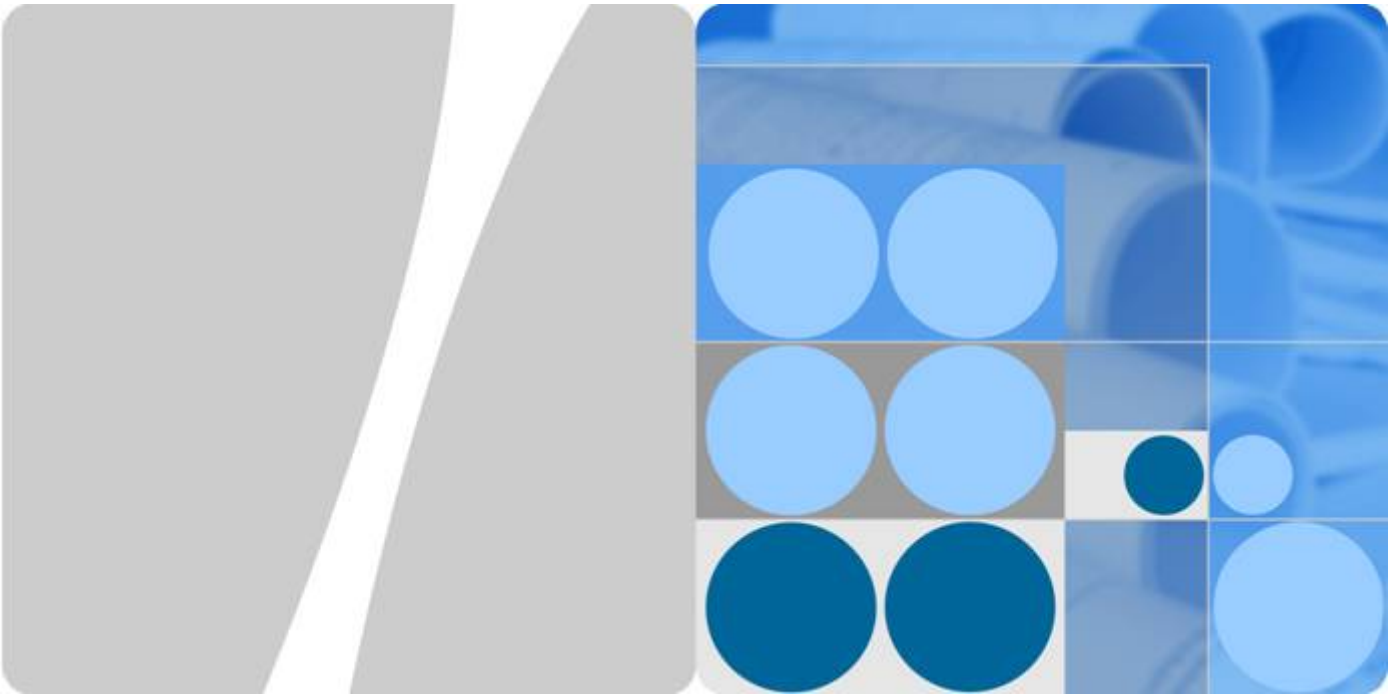


Enterprise Data Communication Products

Feature Description - IP Service

Issue 05

Date 2013-04-25

HUAWEI TECHNOLOGIES CO., LTD.

Copyright © Huawei Technologies Co., Ltd. 2013. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Huawei Technologies Co., Ltd.

Trademarks and Permissions


and other Huawei trademarks are trademarks of Huawei Technologies Co., Ltd.
All other trademarks and trade names mentioned in this document are the property of their respective holders.

Notice

The purchased products, services and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

Huawei Technologies Co., Ltd.

Address: Huawei Industrial Base
Bantian, Longgang
Shenzhen 518129
People's Republic of China

Website: http://enterprise.huawei.com

Copyright © Huawei Technologies Co., Ltd.

About This Document

Intended Audience

This document describes the definition, purpose, and implementation of features on enterprise datacom products including the campus network switch, enterprise router, data center switch, and WLAN. For features supported by the device, see *Configuration Guide*.

This document is intended for:

- Network planning engineers
- Commissioning engineers
- Data configuration engineers
- System maintenance engineers

Symbol Conventions

The symbols that may be found in this document are defined as follows.

Symbol	Description
DANGER	Indicates a hazard with a high level or medium level of risk which, if not avoided, could result in death or serious injury.
WARNING	Indicates a hazard with a low level of risk which, if not avoided, could result in minor or moderate injury.
CAUTION	Indicates a potentially hazardous situation that, if not avoided, could result in equipment damage, data loss, performance deterioration, or unanticipated results.
TIP	Provides a tip that may help you solve a problem or save time.
NOTE	Provides additional information to emphasize or supplement important points in the main text.

Copyright © Huawei Technologies Co., Ltd.

Command Conventions

The command conventions that may be found in this document are defined as follows.

Convention	Description
Boldface	The keywords of a command line are in **boldface**.
Italic	Command arguments are in *italics*.
[ ]	Items (keywords or arguments) in brackets [ ] are optional.
{ x | y | ... }	Optional items are grouped in braces and separated by vertical bars. One item is selected.
[x | y | ...]	Optional items are grouped in brackets and separated by vertical bars. One item is selected or no item is selected.
{ x | y | ... }*	Optional items are grouped in braces and separated by vertical bars. A minimum of one item or a maximum of all items can be selected.
[x | y | ...]*	Optional items are grouped in brackets and separated by vertical bars. You can select one or several items, or select no item.
&<1-n>	The parameter before the & sign can be repeated 1 to n times.
#	A line starting with the # sign is comments.

Change History

Updates between document issues are cumulative. Therefore, the latest document issue contains all updates made in previous issues.

Changes in Issue 05 (2013-04-25)

The second commercial release has the following updates:

The following information is modified:

- **8.2.1 Local PBR**

Changes in Issue 04 (2013-01-31)

The second commercial release has the following updates:

The following information is modified:

- **4.2.3 DHCPv6 Working Principles**

Changes in Issue 03 (2012-12-31)

The second commercial release has the following updates:

Copyright © Huawei Technologies Co., Ltd.

The following information is modified:

- **4.2.1 DHCPv6 Overview**
- **4.2.3 DHCPv6 Working Principles**

Changes in Issue 02 (2012-12-08)

The second commercial release has the following updates:

The documentation is updated according to product feature updates.

Changes in Issue 01 (2012-09-30)

Initial commercial release.

Copyright © Huawei Technologies Co., Ltd.

Contents

About This Document.....ii

1 IPv4.....1

1.1 Introduction to IPv4.....2
1.2 Principles.....2
1.2.1 IPv4 Overview.....2
1.2.2 IPv4 Address.....3
1.2.3 IPv4 Packet Format.....5
1.2.4 Subnetting.....7
1.2.5 IP Address Resolution.....8
1.3 References.....9

2 ARP.....10

2.1 Introduction.....11
2.2 Principles.....11
2.2.1 ARP Principles.....11
2.2.2 Proxy ARP.....15
2.2.3 Gratuitous ARP.....17
2.2.4 ARP-Ping.....17
2.2.5 Layer 2 Proxy ARP.....18
2.2.6 Fast ARP Reply.....19
2.3 References.....20

3 DHCP.....21

3.1 Introduction to DHCP.....22
3.2 Principles.....22
3.2.1 DHCP Overview.....22
3.2.2 Introduction to DHCP Messages.....23
3.2.3 DHCP Options.....26
3.2.4 DHCP Principles.....30
3.2.5 DHCP Relay Principles.....33
3.2.6 IP Address Assignment and Renewal.....35
3.3 Application.....36
3.3.1 DHCP Server Application.....36
3.3.2 DHCP Relay Application.....37

Copyright © Huawei Technologies Co., Ltd.

3.4 References....37
4 DHCPv6....38
4.1 Introduction....39
4.2 Principles....39
4.2.1 DHCPv6 Overview....39
4.2.2 DHCPv6 Packets....42
4.2.3 DHCPv6 Working Principles....45
4.2.4 Working Principle of DHCPv6 PD....47
4.2.5 Working Principle of the DHCPv6 Relay Agent....48
4.2.6 IPv6 Address/Prefix Allocation and Lease Updating....49
4.3 References....52
5 IPv6....53
5.1 Introduction to IPv6....54
5.2 Principles....56
5.2.1 IPv6 Addresses....56
5.2.2 IPv6 Packet Format....62
5.2.3 ICMPv6....66
5.2.4 Neighbor Discovery....68
5.2.5 Path MTU....74
5.2.6 Dual Protocol Stack....75
5.2.7 IPv6 over IPv4 Tunnel....76
5.2.8 IPv4 over IPv6 Tunnel....84
5.3 References....85
6 DNS....87
6.1 Introduction to DNS....88
6.2 Principles....88
6.2.1 Working Principle of DNS....88
6.2.2 Working Principle of DNS Proxy or Relay....90
6.2.3 Working Principle of DNS Spoofing....91
6.2.4 Working Principle of DDNS....93
6.3 Applications....94
6.3.1 DNS Client Application....94
6.3.2 DNS Proxy Application....94
6.4 References....95
7 NAT....96
7.1 Introduction to NAT....97
7.2 Principles....97
7.2.1 Overview....97
7.2.2 NAT Implementation....99
7.2.3 NAT ALG....102
7.2.4 DNS Mapping....103

Copyright © Huawei Technologies Co., Ltd.

7.2.5 NAT Associated with VPNs....104
7.2.6 Twice NAT....106
7.2.7 NAT Filtering and NAT Mapping....107
7.3 Applications....109
7.3.1 Private Network Hosts Accessing Public Network Servers....109
7.3.2 Public Network Hosts Accessing Private Network Servers....110
7.3.3 Private Network Hosts Accessing Private Network Servers Using the Domain Name....110
7.3.4 NAT Multi-instance....111
7.4 References....112
8 IP Unicast Policy-based Routing....114
8.1 Introduction to IP Unicast Policy-based Routing....115
8.2 Principles....115
8.2.1 Local PBR....115
8.2.2 Interface PBR....117
8.2.3 Smart Policy Routing....117
8.3 Applications....120
8.4 References....122

Copyright © Huawei Technologies Co., Ltd.

1 IPv4

About This Chapter

1.1 Introduction to IPv4

1.2 Principles

1.3 References

Copyright © Huawei Technologies Co., Ltd.

1.1 Introduction to IPv4

Definition

Internet Protocol Version 4 (IPv4) is the core protocol in the TCP/IP protocol suite. IPv4 works at the network layer in the TCP/IP model. This layer corresponds to the network layer in the Open System Interconnection Reference Model (OSI RM). The network layer provides connectionless data transmission. Each IP datagram is transmitted independently.

Purpose

IPv4 is used on the network layer between the data link layer and the transport layer. IPv4 shields the differences at the link layer and provides a uniform format for the data packets transmitted at the transport layer.

1.2 Principles

1.2.1 IPv4 Overview

IPv4 Protocol Suite

Internet Protocol Version 4 (IPv4) is the core protocol in the TCP/IP protocol suite. IPv4 protocol suite includes Address Resolution Protocol (ARP), Reverse Address Resolution Protocol (RARP), Internet Control Message Protocol (ICMP), Transmission Control Protocol (TCP), and User Datagram Protocol (UDP).

Figure 1-1 IPv4 protocol suite

Transport layer	TCP，UDP
Network layer	ICMP IP RARP，ARP
Data link layer	Various network interfaces

As shown in **Figure 1-1**, ARP and RARP work between the data link layer and the network layer for address resolution. ICMP works between the network layer and the transport layer to ensure correct forwarding of IP datagrams.

ARP

ARP maps an IP address to a MAC address. ARP can be implemented in dynamic or static mode. ARP provides some extended functions, such as proxy ARP, gratuitous ARP, ARP security, and ARP-Ping.

Copyright © Huawei Technologies Co., Ltd.

RARP

RARP maps a MAC address to an IP address.

ICMP

ICMP works at the network layer to ensure correct forwarding of IP datagrams. ICMP allows hosts and routers to report errors during packet transmission. An ICMP message is encapsulated in an IP datagram as the data, and a header is added to the ICMP message to form an IP datagram.

1.2.2 IPv4 Address

To connect a PC to the Internet, you need to apply an IP address from the Internet Service Provider (ISP).

An IP address is a numerical label assigned to each device on a computer network. An IPv4 address is a 32-bit binary number. IP addresses are expressed in dotted decimal notation, which helps you memorize and identify them. In dotted decimal notation, an IPv4 address is written as four decimal numbers, one for each byte of the address. For example, the binary IPv4 address 00001010 00000001 00000001 00000010 is written as 10.1.1.2 in dotted decimal notation.

An IPv4 address consists of two parts:

- Network ID (Net-id).The network ID identifies a network. The leftmost several bits of the network ID identify the class of IP addresses.
- Host ID (Host-id).The host ID identifies different hosts on a network. Network devices with the same network ID are located on the same network, regardless of their physical locations.

Characteristics of IPv4 Addresses

IPv4 addresses have the following characteristics:

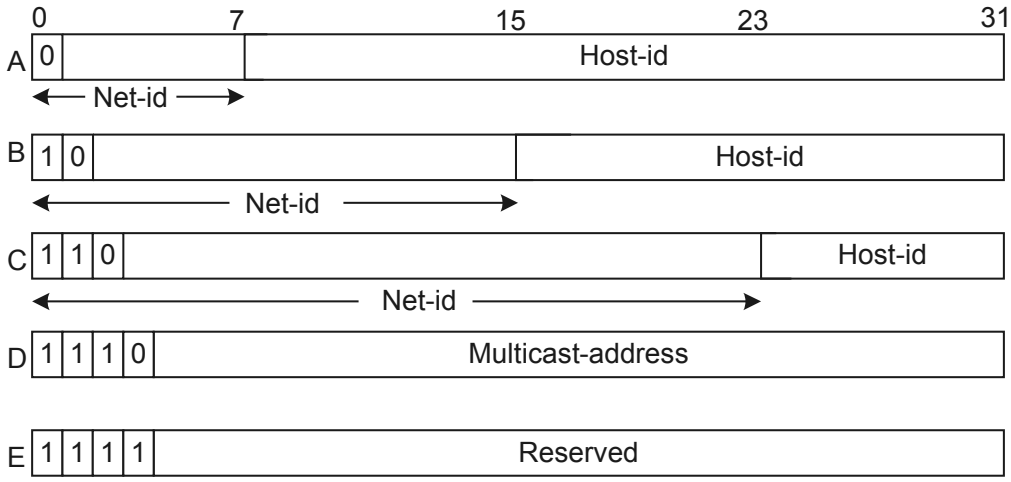
- IP addresses do not show any geographical information. The network ID represents the network to which a host belongs.
- When a host connects to two networks simultaneously, it must have two IP addresses with different network IDs. In this case, the host is called a multihomed host. Each interface on a host has an IP address. Therefore, a multi-interface host has multiple IP addresses.
- Networks allocated with the network ID are in the same class.

IPv4 Address Classification

As shown in **Figure 1-2**, IP addresses are classified into five classes to facilitate IP address management and networking.

Copyright © Huawei Technologies Co., Ltd.

Figure 1-2 Five classes of IP addresses



At present, most IP addresses in use belong to Class A, Class B, or Class C. Class D addresses are multicast addresses and Class E addresses are reserved. The easiest way to determine the class of an IP address is to check the first bits in its network ID. The class fields of Class A, Class B, Class C, Class D, and Class E are binary digits 0, 10, 110, 1110, and 1111 respectively. For details about IP address classification, see RFC 1166 (Internet Numbers).

Certain IP addresses are reserved, and they cannot be allocated to users. **Table 1-1** lists the ranges of IP addresses for the five classes.

Table 1-1 IP address classes and ranges

Class	Range	Description
A	0.0.0.0 to 127.255.255.255	IP addresses with all-0 host IDs are network addresses and are used for network routing. IP addresses with all-1 host IDs are broadcast addresses and are used for broadcasting packets to all hosts on the network.
B	128.0.0.0 to 191.255.255.255	IP addresses with all-0 host IDs are network addresses and are used for network routing. IP addresses with all-1 host IDs are broadcast addresses and are used for broadcasting packets to all hosts on the network.
C	192.0.0.0 to 223.255.255.255	IP addresses with all-0 host IDs are network addresses and are used for network routing. IP addresses with all-1 host IDs are broadcast addresses and are used for broadcasting packets to all hosts on the network.
D	224.0.0.0 to 239.255.255.255	Class D addresses are multicast addresses.
E	240.0.0.0 to 255.255.255.255	Reserved. The IP address 255.255.255.255 is used as a Local Area Network (LAN) broadcast address.

Copyright © Huawei Technologies Co., Ltd.

Special IPv4 Addresses

Table 1-2 Special IP addresses

Network ID	Host ID	Used as a Source Address	Used as a Destination Address	Description
All 0s	All 0s	Yes	No	Used by local hosts on a local network.
All 0s	Host ID	Yes	No	Used by specified hosts on a network.
127	Any value except all 0s or all 1s	Yes	Yes	Used as loopback addresses.
All 1s	All 1s	No	Yes	Limited broadcast address (packets with this IP address will never be forwarded).
Net-id	All 1s	No	Yes	Directed broadcast address (packets with this IP address is broadcast on the specified network).

NOTE

Net-id is neither all 0s nor all 1s.

Private IPv4 Addresses

Private IP addresses are used to solve the problem of IP address shortage. Private addresses are used on internal networks or hosts, and cannot be used on the public network. RFC 1918 describes three IP address segments reserved for private networks.

Table 1-3 Private IP addresses

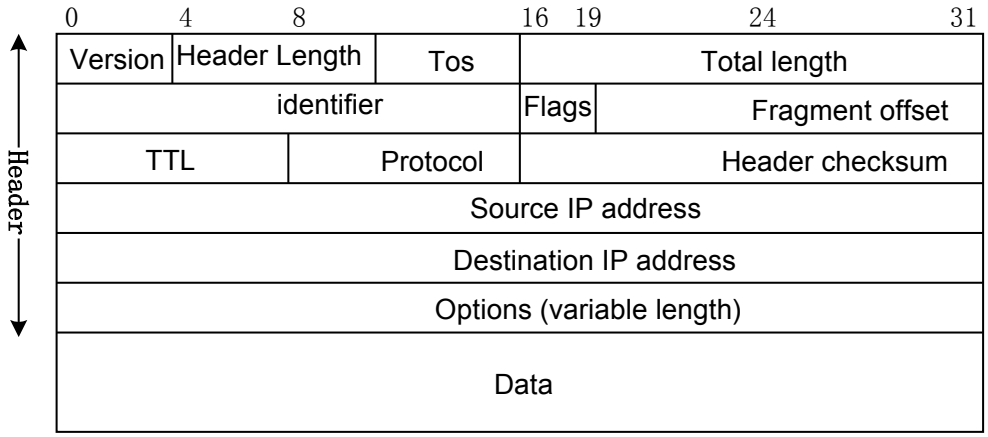
Class	Range
A	10.0.0.0 to 10.255.255.255
B	172.16.0.0 to 172.31.255.255
C	192.168.0.0 to 192.168.255.255

1.2.3 IPv4 Packet Format

Figure 1-3 shows the IPv4 packet format.

Copyright © Huawei Technologies Co., Ltd.

Figure 1-3 IPv4 packet format



An IPv4 datagram consists of a header and a data field. The first 20 bytes in the header are mandatory for all IPv4 datagrams. The Options field following the 20 bytes has a variable length.

Table 1-4 describes the meaning of each field in an IPv4 packet.

Table 1-4 Description of each field in an IPv4 packet

Field	Length	Description
Version	4 bits	Specifies the IP protocol version, IPv4 or IPv6.
Header Length	4 bits	Specifies the length of the IPv4 header.
Type of Service (ToS)	8 bits	Specifies the type of service. This field takes effect only in the differentiated service model.
Total Length	16 bits	Specifies the length of the header and data.
Identification	16 bits	IPv4 software maintains a counter in the storage device to record the number of IP datagrams. The counter value increases by 1 every time a datagram is sent, and is filled in the identification field.
Flags	3 bits	Only the rightmost two bits are valid. The rightmost bit indicates whether the datagram is not the last data fragment. The value 1 indicates the last fragment, and the value 0 indicates non-last fragment. The middle bit is the fragmentation flag. The value 1 indicates that the datagram cannot be fragmented, and the value 0 indicates that the datagram can be fragmented.
Fragment Offset	13 bits	Specifies the location of a fragment in a packet.
Time to Live (TTL)	8 bits	Specifies the life span of a datagram on a network. TTL is measured by the number of hops.
Protocol	8 bits	Specifies the type of the protocol carried in the datagram.

Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
Header Checksum	16 bits	A router calculates the header checksum for each datagram received. If the checksum is 0, the router knows that the header remains unchanged and retains the datagram. This field checks only the header but not the data.
Source IP Address	32 bits	Specifies the IPv4 address of a sender.
Destination IP Address	32 bits	Specifies the IPv4 address of a receiver.
Options (variable length)	0-32 bits	Allows IPv4 to support various options such as fault handling, measurement, and security. The Options field always ends on a 32-bit boundary. Pad bytes with a value of 0 are added if necessary.
Data	Variable	Pads an IP datagram .

1.2.4 Subnetting

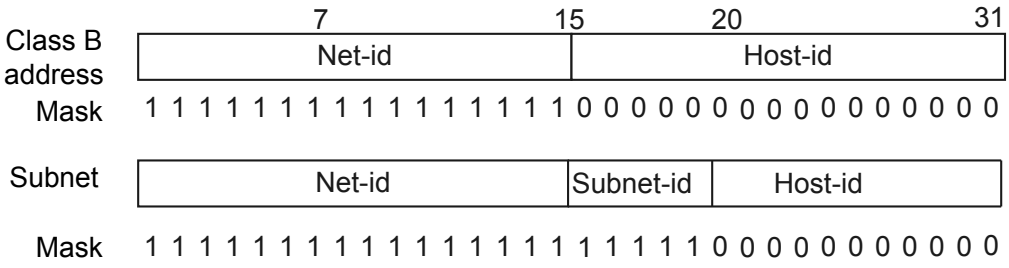
A network can be divided into multiple subnets to conserve IP address space and support flexible IP addressing.

When many hosts are distributed on an internal network, the internal host IDs can be divided into multiple subnet IDs to facilitate management. Then the entire network contains multiple small networks.

Subnetting is implemented within the internal network. The internal network has only one network ID for the external network. When packets are transmitted from the external network to the internal network, the router on the internal network selects a route for the packets based on the subnet ID and finds the destination host.

Figure 1-4 shows subnetting of a Class B IP address. The subnet mask consists of a string of continuous 1s and 0s. 1s indicate the network ID and the subnet ID field, and 0s indicate the host ID.

Figure 1-4 Subnetting of IP addresses



As shown in **Figure 1-4**, the first 5 bits of the host ID is used as the subnet ID. The subnet ID ranges from 00000 to 11111, allowing a maximum of 32 (2^5) subnets. Each subnet ID has a

Copyright © Huawei Technologies Co., Ltd.

subnet mask. For example, the subnet mask of the subnet ID 11111 is 255.255.248.0. After performing an AND operation on the IP address and the subnet mask, you can obtain the network address.

Subnetting reduces the available IP addresses. For example, a Class B IP address contains 65534 host IDs. After 5 bits in the host ID are used as the subnet ID, there can be a maximum of 32 subnets, each having an 11-bit host ID. Each subnet has a maximum of 2046 host IDs (2^{11} - 2, excluding the host IDs with all 1s and all 0s). Therefore, the IP address has a maximum of 65472 (32 x 2046) host IDs, 62 less than the maximum number of host IDs before subnetting.

To implement efficient network planning, subnetting and IP addressing should abide by the following rules.

Hierarchy

To divide a network into multiple layers, you need to consider geographic and service factors. Use a top-down subnetting mode to facilitate network management and simplify routing tables. In most cases:

- A network consisting of a backbone network and a MAN is divided into hierarchical subnets.
- An administrative network is divided into subnets based on administrative levels.

Consecutiveness

Consecutive addresses facilitate route summarization on a hierarchical network, which greatly reduces the number of routing entries and improves route search efficiency.

- Allocate consecutive IP addresses to each area.
- Allocate consecutive IP addresses to devices that have the same services and functions.

Scalability

When allocating addresses, reserve certain addresses on each layer to ensure consecutive address allocation in future network expansion.

A backbone network must have enough consecutive addresses for independent autonomous systems (ASs) and further network expansion.

Efficiency

When planning subnets, fully utilize address resources to ensure that the subnets are sufficient for hosts.

- Allocate IP addresses by using variable-length subnet masking (VLSM) to fully use address resources.
- Consider the routing mechanisms in IP address planning to improve address utilization efficiency in the allocated address spaces.

1.2.5 IP Address Resolution

A router that connects to multiple networks has the IP addresses of the connected networks. To ensure that users can use the IP address normally, ensure that:

Copyright © Huawei Technologies Co., Ltd.

- An IP address is a network layer address of a host. To transmit data packets to a destination host, the router must obtain the physical address of the host. Therefore, the IP address must be resolved to a physical address.
- A host name is easier to remember than an IP address. Therefore, the host name needs to be resolved to the IP address.

On Ethernet, the physical address of a host is the MAC address. The DNS server resolves a host name to an IP address. ARP resolves an IP address to a MAC address. For details, see **DNS** and **ARP**.

1.3 References

The following table lists the references of the IPv4 feature.

Document	Description	Remarks
RFC1166	Internet Numbers	-
RFC1918	Address Allocation for Private Internets	-

Copyright © Huawei Technologies Co., Ltd.

2 ARP

About This Chapter

2.1 Introduction

2.2 Principles

2.3 References

Copyright © Huawei Technologies Co., Ltd.

2.1 Introduction

Definition

The Address Resolution Protocol (ARP) maps IP addresses into MAC addresses.

Purpose

On a local area network (LAN), a host or a network device must learn the IP address of the destination host or device before sending data to it. Additionally, the host or network device must learn the physical address of the destination host or device because IP packets must be encapsulated into frames for transmission over a physical network. Therefore, the mapping from an IP address into a physical address is required. ARP is used to map IP addresses into physical addresses.

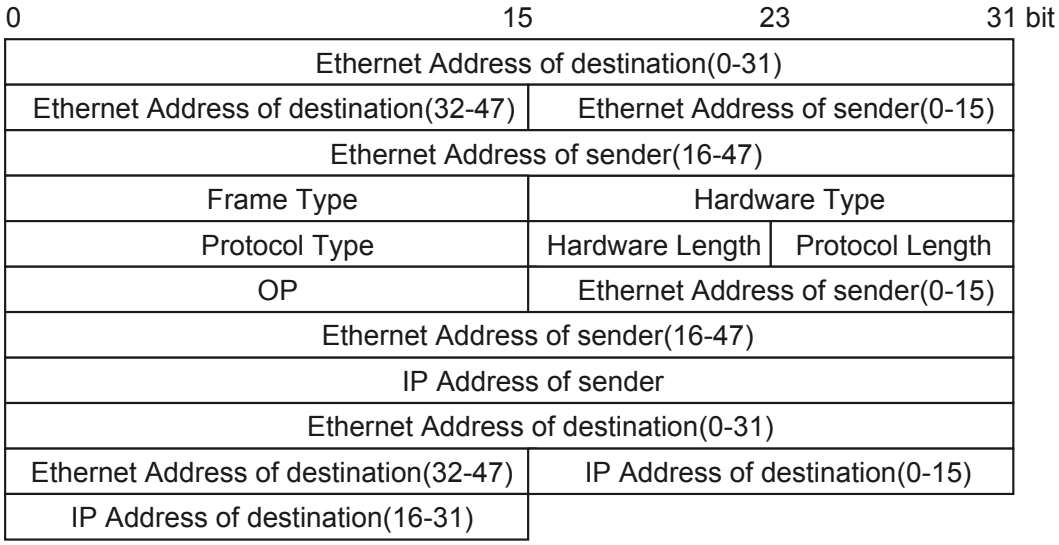
2.2 Principles

2.2.1 ARP Principles

Format of ARP Packets

Figure 2-1 shows the format of an ARP Request or Reply packet.

Figure 2-1 Format of an ARP Request or Reply packet



Description of the main fields is as follows:

- Hardware Type: indicates the hardware address type. For an Ethernet, the value of this field is 1.

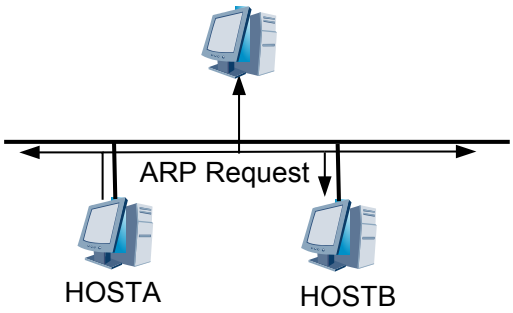
Copyright © Huawei Technologies Co., Ltd.

- Protocol Type: indicates the type of the protocol address to be mapped. For an IP address, the value of this field is 0x0800.
- Hardware Length: indicates the hardware address length. For an ARP Request or Reply packet, the value of this field is 6.
- Protocol Length: indicates the protocol address length. For an ARP Request or Reply packet, the value of this field is 4.
- OP: indicates the operation type. The value 1 indicates ARP requesting, and the value 2 indicates ARP replying.
- Ethernet Address of sender: indicates the MAC address of the sender.
- IP Address of sender: indicates the IP address of the sender.
- Ethernet Address of destination: indicates the MAC address of the receiver.
- IP Address of destination: indicates the IP address of the receiver.

Address Resolution Process

ARP completes address resolution through two processes: ARP request process and ARP reply process.

Figure 2-2 ARP request process



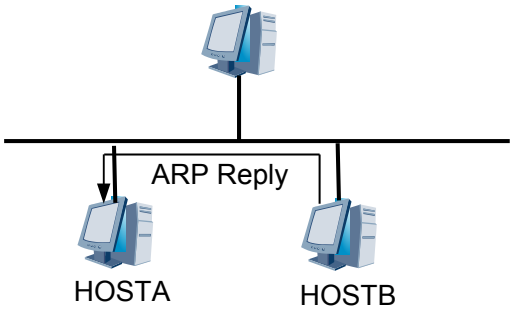
As shown in **Figure 2-2**, HOSTA and HOSTB are on the same network segment. HOSTA needs to send IP packets to HOSTB.

HOSTA searches the local ARP table for the ARP entry corresponding to HOSTB. If the corresponding ARP entry is found, HOSTA encapsulates the IP packets into Ethernet frames and forwards them to HOSTB based on its MAC address.

If the corresponding APR entry is not found, HOSTA caches the IP packets and broadcasts an ARP Request packet. In the ARP Request packet, the IP address and MAC address of the sender are the IP address and MAC address of HOSTA. The destination IP address is the IP address of HOSTB, and the destination MAC address contains all 0s. All hosts on the same network segment can receive the ARP Request packet, but only HOSTB processes the packet.

Copyright © Huawei Technologies Co., Ltd.

Figure 2-3 ARP reply process



HOSTB compares its IP address with the destination IP address in the ARP Request packet. If HOSTB finds that its IP address is the same as the destination IP address, HOSTB adds the IP address and MAC address of the sender (HOSTA) to the local ARP table. Then HOSTB unicasts an ARP Reply packet, which contains its MAC address, to HOSTA, as shown in **Figure 2-3**.

After receiving the ARP Reply packet, HOSTA adds HOSTB's MAC address into the local ARP table. Meanwhile, HOSTA encapsulates the IP packets and forwards them to HOSTB.

NOTE

The MAC address that hostA learned may be a multicast MAC address.

ARP Aging Mechanism

- ARP cache (ARP table)

  If HOSTA broadcasts an ARP Request packet every time it communicates with HOSTB, the communication traffic on the network will increase. Furthermore, all hosts on the network have to receive and process the ARP Request packet, which decreases network efficiency.

  To solve the preceding problems, each host maintains an ARP cache, which is the key to efficient operation of ARP. This cache contains the recent mapping from IP addresses to MAC addresses.

  Before sending IP packets, a host searches the cache for the MAC address corresponding to the destination IP address. If the cache contains the MAC address, the host does not send an ARP Request packet but directly sends the IP packets to the destination MAC address. If the cache does not contain the MAC address, the host broadcasts an ARP Request packet on the network.

- Aging time of dynamic ARP entries

  After HOSTA receives the ARP Reply packet from HOSTB, HOSTA adds the mapping between the IP address and the MAC address of HOSTB to the ARP cache. However, if a fault occurs on HOSTB or the network adapter of HOSTB is replaced but HOSTA is not notified, HOSTA still sends IP packets to HOSTB. This fault occurs because the APR entry of HOSTB in the ARP cache on HOSTA is not updated.

  To reduce address resolution errors, a timer is set for each ARP entry in an ARP cache. A dynamic APR entry is deleted when its timer expires.

  Configuring the timer reduces address resolution errors but does not eliminate the problem because of the time delay. Specifically, if the length of a dynamic APR entry timer is N

Copyright © Huawei Technologies Co., Ltd.

seconds, the sender can detect the fault on the receiver after N seconds. During the N seconds, the cache on the sender is not updated.

- Number of probes for aging dynamic ARP entries

  Besides setting a timer for dynamic ARP entries, you can set the number of probes for aging dynamic ARP entries to reduce address resolution errors. Before aging a dynamic ARP entry, a host sends ARP aging probe packets. If the host receives no ARP Reply packet after the number of probes reaches the maximum number, the ARP entry is deleted.

- Aging probe modes for dynamic ARP entries

  Before a dynamic ARP entry on a device is aged out, the device sends ARP aging probe packets to other devices on the same network segment. An ARP aging probe packet can be a unicast or broadcast packet. By default, a device broadcasts ARP aging probe packets.

  If the IP address of the peer device remains the same but the MAC address changes frequently, it is recommended that you configure ARP aging probe packets to be broadcast.

  If the MAC address of the peer device remains the same, the network bandwidth is insufficient, and the aging time of ARP entries is short, it is recommended that you configure ARP aging probe packets to be unicast.

  When a non-Huawei device connected to a Huawei device receives an ARP aging probe packet whose destination MAC address is a broadcast address, the non-Huawei device checks the ARP table. If the mapping between the IP address and the MAC address of the Huawei device exists in the ARP table, the non-Huawei device drops the ARP aging probe packet. The Huawei device cannot receive a response and therefore deletes the corresponding ARP entry. As a result, traffic from the network cannot be forwarded. In this scenario, the Huawei device needs to send ARP aging probe packets in unicast mode and the non-Huawei device needs to respond to the ARP aging probe packets.

- Layer 2 topology detection

  The Layer 2 topology detection function enables a device to retransmit ARP probe packets to update ARP entries when a Layer 2 interface becomes Up and the aging time of the ARP entries in the corresponding VLAN becomes 0.

Dynamic ARP

Dynamic ARP entries are generated and maintained dynamically by using ARP packets. They can be aged out, updated, or overwritten by static ARP entries. When the aging time expires or the interface is Down, the corresponding dynamic ARP entries are deleted.

Static ARP

Static ARP entries record fixed mapping between IP addresses and MAC addresses and are configured manually by network administrators. Devices cannot dynamically change the mapping.

Static ARP is configured to:

- Forward packets whose destination IP addresses are not on the local network segment to a gateway on the local network segment. Then the packets can be forwarded by the gateway.
- Bind the destination IP addresses of invalid packets to a nonexistent MAC address so that the invalid packets are filtered out.

Copyright © Huawei Technologies Co., Ltd.

2.2.2 Proxy ARP

If an ARP Request packet is sent to a host on a different network, the routing device that connects the two networks can reply to this packet. This function is called proxy ARP.

Proxy ARP has the following characteristics:

- Proxy ARP is implemented on the ARP subnet gateway without any modifications on any hosts.
- Proxy ARP can shield topologies of physical networks so that hosts on different physical networks can use the same network ID to communicate. Proxy ARP enables hosts that are on the same network segment but on different physical networks to communicate.
- Proxy ARP affects only the ARP caches on hosts but does not affect the ARP cache or routing table on the gateway.
- After proxy ARP is enabled, the aging time of ARP entries on hosts should be shortened so that invalid ARP entries can be deleted as soon as possible. Then IP packet forwarding failures decrease on the router.

The following table shows three types of proxy ARP.

Proxy ARP Type	Resolved Issue
Routed proxy ARP	Allows hosts on the same network segment but on different physical networks to communicate.
Intra-VLAN proxy ARP	Allows isolated hosts in a VLAN to communicate.
Inter-VLAN proxy ARP	Allows hosts in different VLANs or hosts in different sub-VLANs of the same VLAN to communicate at Layer 3.

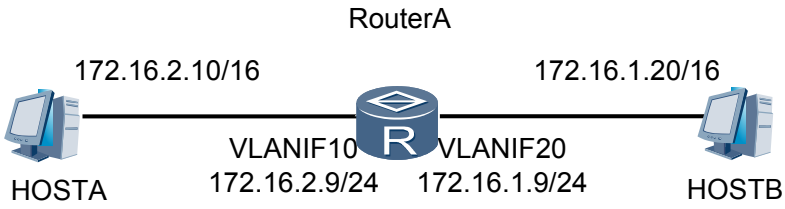
Routed Proxy ARP

Routed proxy ARP enables network devices on the same network segment but on different physical networks to communicate.

In practice, if a host connected to a router is not configured with a default gateway address (that is, the host does not know how to reach the intermediate system of the network), the host cannot transmit packets.

As shown in **Figure 2-4**, RouterA is connected to two networks through VLAN10 and VLAN20. The IP addresses of VLANIF10 and VLANIF20 are on different network segments. However, the masks make HOSTA and VLANIF10 on the same network segment, HOSTB and VLANIF20 on the same network segment, and HOSTA and HOSTB on the same network segment.

Figure 2-4 Application of routed proxy ARP



Copyright © Huawei Technologies Co., Ltd.

The IP addresses of HOSTA and HOSTB are on the same network segment. When HOSTA needs to communicate with HOSTB, HOSTA broadcasts an ARP Request packet, requesting the MAC address of HOSTB. However, HOSTA and HOSTB are on different physical networks (in different broadcast domains). Therefore, HOSTB cannot receive the ARP Request packet sent from HOSTA and does not respond with an ARP Reply packet.

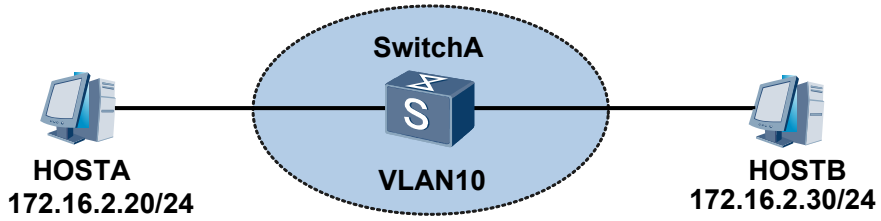
To solve this problem, enable proxy ARP on RouterA. After receiving an ARP Request packet, RouterA enabled with proxy ARP searches for the routing table corresponding to HOSTB. If the route corresponding to HOSTB exists, RouterA responds to the ARP Request packet with its own MAC address. HOSTA forwards data based on the MAC address of RouterA. RouterA functions as the proxy of HOSTB.

Intra-VLAN Proxy ARP

If two hosts belong to the same VLAN but are isolated, enable intra-VLAN proxy ARP on an interface associated with the VLAN to allow the hosts to communicate.

As shown in **Figure 2-5**, HOSTA and HOSTB are connected to SwitchA. The two interfaces connected to HOSTA and HOSTB belong to VLAN10.

Figure 2-5 Application of intra-VLAN proxy ARP



HOSTA and HOSTB cannot communicate at Layer 2 because interface isolation in a VLAN is configured on SwitchA.

To solve this problem, enable intra-VLAN proxy ARP on the interfaces of SwitchA. After SwitchA's interface connected to HOSTA receives an ARP Request packet whose destination address is not its own address, SwitchA does not discard the packet but searches for the ARP entry corresponding to HOSTB. If the ARP entry corresponding to HOSTB exists, SwitchA sends its MAC address to HOSTA and forwards packets sent from HOSTA to HOSTB. SwitchA functions as the proxy of HOSTB.

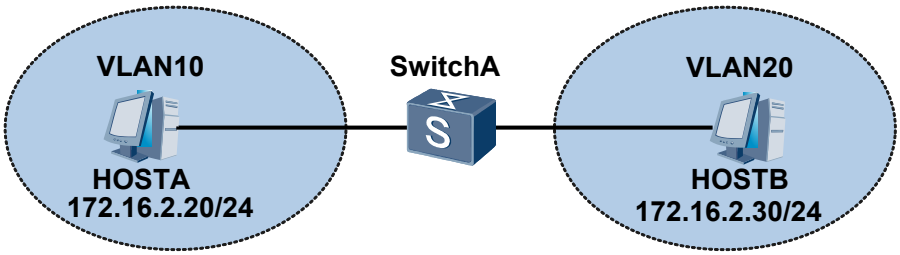
Inter-VLAN Proxy ARP

If two hosts belong to different VLANs, enable inter-VLAN proxy ARP on interfaces associated with the VLANs to implement Layer 3 communication between the two hosts.

As shown in **Figure 2-6**, HOSTA and HOSTB are connected to SwitchA. The interface connected to HOSTA belongs to VLAN10, and the interface connected to HOSTB belongs to VLAN20.

Copyright © Huawei Technologies Co., Ltd.

Figure 2-6 Application of inter-VLAN proxy ARP



The interfaces connected to HOSTA and HOSTB belong to different VLANs. Therefore, HOST A and HOSTB cannot communicate at Layer 2.

To solve this problem, enable inter-VLAN proxy ARP on the interfaces of SwitchA. After SwitchA's interface connected to HOSTA receives an ARP Request packet whose destination address is not its own address, SwitchA does not discard the packet but searches for the ARP entry corresponding to HOSTB. If the ARP entry corresponding to HOSTB exists, SwitchA sends its MAC address to HOSTA and forwards packets sent from HOSTA to HOSTB. SwitchA functions as the proxy of HOSTB.

2.2.3 Gratuitous ARP

Gratuitous ARP enables a host to send an ARP Request packet using its own IP address as the destination address. Gratuitous ARP provides the following functions:

- Checks duplicate IP addresses: Normally, a host does not receive an ARP Reply packet after sending an ARP Request packet with the destination address being its own IP address. If the host receives an ARP Reply packet, another host has the same IP address.
- Advertises a new MAC address. If the MAC address of a host changes because its network adapter is replaced, the host sends a gratuitous ARP packet to notify all hosts of the change before the ARP entry is aged out.
- Notifies an active/standby switchover in a VRRP backup group: After an active/standby switchover, the master router sends a gratuitous ARP packet in the VRRP backup group to notify the switchover.

2.2.4 ARP-Ping

ARP-Ping includes ARP-Ping IP and ARP-Ping MAC. ARP-Ping sends ARP Request packets or ICMP Echo Request packets to check whether a specified IP address or MAC address is used.

ARP-Ping IP

ARP-Ping IP checks whether an IP address is used by another device on the LAN by sending ARP packets.

Before configuring an IP address for a device, configure ARP-Ping IP on the device to check whether this IP address has been used by sending ARP Request packets.

You can also run the ping command to check whether this IP address is used by another device on the network. However, if the routing device or host that uses the IP address is enabled with the firewall function and the firewall is configured not to respond to ping packets, you may be misled into thinking that this IP address is not used. To solve the problem, use ARP-Ping IP.

Copyright © Huawei Technologies Co., Ltd.

ARP is a Layer 2 protocol. Therefore, ARP packets can pass through the firewall that is configured not to respond to ping packets.

ARP-Ping IP sends ARP Request packets. ARP-Ping IP is implemented as follows:

1. After an IP address is specified for a host using the **arp-ping ip** command, the host sends an ARP Request packet and starts a timer of waiting for an ARP Reply packet.
2. After receiving the ARP Request packet, the routing device or host that uses this IP address in the LAN returns an ARP Reply packet.
3. The sender performs the following two operations based on whether it receives the ARP packet:
   - If the sender receives an ARP Reply packet, the sender compares the source IP address carried in the ARP Reply packet with the IP address specified using the **arp-ping ip** command. If the two IP addresses are the same, the MAC address corresponding to the specified IP address is displayed and the timer is disabled.
   - If the sender does not receive an ARP Reply packet before the timer of waiting for an ARP Reply packet expires, the sender displays a message indicating that the IP address is not used by another routing device or host.

ARP-Ping MAC

The ARP-Ping MAC process is similar to the ping process. The difference is that ARP-Ping MAC applies only to directly connected Ethernet LANs or Layer 2 VPN Ethernet networks.

ARP-Ping MAC sends ICMP Echo Request packets. ARP-Ping MAC is implemented as follows:

1. After a MAC address is specified for a host using the **arp-ping mac** command, the host sends an ICMP Echo Request packet and starts a timer of waiting for an ICMP Echo Reply packet.
2. After receiving the ICMP Echo Request packet, the routing device or host that uses this MAC address in the LAN returns an ICMP Echo Reply packet.
3. The sender performs the following two operations based on whether it receives the ICMP packet:
   - If the sender receives an ICMP Echo Reply packet, the sender compares the source MAC address carried in the ICMP Echo Reply packet with the MAC address specified using the **arp-ping mac** command. If the two MAC addresses are the same, the sender displays the source IP address of the ICMP Echo Reply packet and displays a message indicating that the MAC address is used by another routing device or host. The timer is disabled.
   - If the sender does not receive an ARP Reply packet before the timer of waiting for an ICMP Echo Reply packet expires, the sender displays a message indicating that the MAC address is not used by another routing device or host.

2.2.5 Layer 2 Proxy ARP

Possible Causes

ARP request messages are broadcast. When receiving an ARP request message, the switching device broadcasts the message within its broadcast domain. If a switching device receives a large number of ARP request messages and broadcasts them, excessive network resources are consumed. The network is congested and the performance deteriorates. Therefore, services are affected.

Copyright © Huawei Technologies Co., Ltd.

Layer 2 proxy ARP effectively distributes the pressure of processing ARP messages by isolating ARP broadcast domains and proxy responding to ARP request messages with local messages. Layer 2 proxy ARP applies to access or convergence devices (such as TOR/EOR in a data center) that connect the gateway and the user.

Implementation Process

When receiving ARP request messages, switching devices check destination IP addresses of the messages.

- If the destination IP address is that of a VLANIF interface, ARP reply messages are processed normally.
- If the destination IP address is not that of a VLANIF interface, the device queries DHCP snooping and ARP snooping entries based on the destination IP address.

  1. Query DHCP snooping entries.
     - If the query succeeds, the device responds to ARP request messages carrying information of the DHCP snooping entries.
     - If the query fails, go to step 2.

     NOTE

     For details on DHCP Snooping, see DHCP Snooping in the *Feature Description - Security*.
  2. Query ARP snooping entries.
     - If the query succeeds, the device responds to ARP request messages carrying information of the ARP snooping entries.
     - If the query fails, the device processes the ARP request messages based on the original procedure.

     NOTE

     ARP snooping is a feature applied to Layer 2 switching networks. Devices use ARP snooping to monitor ARP messages and sets up ARP snooping entries recording user information. The information includes source IP addresses of ARP messages, source MAC addresses, inbound interfaces of the messages, and VLANs that the interfaces belong to.

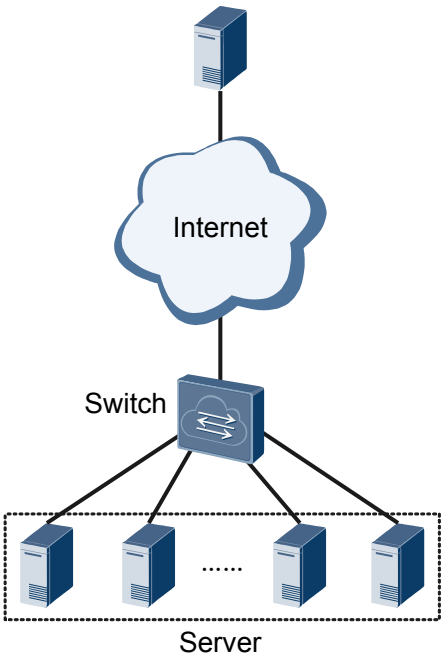
2.2.6 Fast ARP Reply

As shown in **Figure 2-7**, the Switch functioning as a gateway connects to multiple servers or virtual machines (VMs). The servers or VMs send ARP Request packets to the gateway at a fixed interval to detect whether the gateway is working properly. The destination IP address of the ARP Request packets is the IP address of the gateway. Processing a large number of ARP Request packets slows the ARP response of the gateway and even causes the gateway unable to respond to the ARP Request packets of some servers or VMs. The ARP entries of the servers or VMs are then aged out on the gateway. As a result, packet loss or service interruption occurs.

Fast ARP reply can reduce load on the gateway and improve ARP packet processing efficiency. When the gateway learns the ARP entries of the servers or VMs, the gateway directly sends ARP Response packets to the servers or VMs without learning ARP entries again. This method reduces the packet processing pressure on the gateway. Fast ARP response allows a device to directly send ARP Response packets in response to ARP Request packets with the destination IP address as the IP address of the device. This function speeds up ARP response and ensures uninterrupted service forwarding.

Copyright © Huawei Technologies Co., Ltd.

Figure 2-7 Typical networking for fast ARP response



2.3 References

The following table lists the references of this document.

Document	Description	Remarks
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	-

Copyright © Huawei Technologies Co., Ltd.

3 DHCP

About This Chapter

3.1 Introduction to DHCP

3.2 Principles

3.3 Application

3.4 References

Copyright © Huawei Technologies Co., Ltd.

3.1 Introduction to DHCP

Definition

The Dynamic Host Configuration Protocol (DHCP) dynamically assigns IP addresses to users and manages user configurations in a centralized manner.

Purpose

As the network expands and becomes complex, the number of hosts often exceeds the number of available IP addresses. As portable computers and wireless networks are widely used, the positions of computers often change, causing IP addresses of the computers to be changed accordingly. As a result, network configurations become increasingly complex. To properly and dynamically assign IP addresses to hosts, DHCP is used.

DHCP is developed based on the BOOTstrap Protocol (BOOTP). BOOTP runs on networks where each host has a fixed network connection. The administrator configures a BOOTP parameter file for each host, and the file remains unchanged for a long period of time. DHCP has the following new features compared with BOOTP:

- Dynamically assigns IP addresses and configuration parameters to clients.
- Enables a host to obtain an IP address dynamically, but does not specify an IP address for each host.

DHCP rapidly and dynamically allocates IP addresses, which improves IP address usage.

3.2 Principles

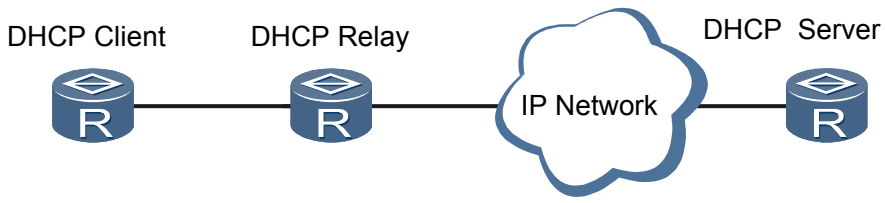
3.2.1 DHCP Overview

DHCP uses the client/server model. A DHCP client sends a packet to a DHCP server to request configuration parameters such as the IP address, subnet mask, and default gateway address. The DHCP server responds with a packet carrying the requested configurations based on a policy.

DHCP Architecture

Figure 3-1 shows the DHCP architecture.

Figure 3-1 DHCP architecture



DHCP involves the following roles:

Copyright © Huawei Technologies Co., Ltd.

- DHCP Client

  A DHCP client exchanges messages with a DHCP server to obtain an IP address and other configuration parameters. On the device, an interface can function as a DHCP client to dynamically obtain configuration parameters such as an IP address from a DHCP server. This facilitates configurations and centralized management.

- DHCP Relay

  A DHCP relay agent forwards DHCP packets exchanged between a DHCP client and a DHCP server that are located on different network segments so that they can complete their address configuration. Using a DHCP relay agent eliminates the need for deploying a DHCP server on each network segment. This feature reduces network deployment costs and facilitates device management.

  In the DHCP architecture, the DHCP relay agent is optional. A DHCP relay agent is required only when the server and client are located on different network segments.

- DHCP Server

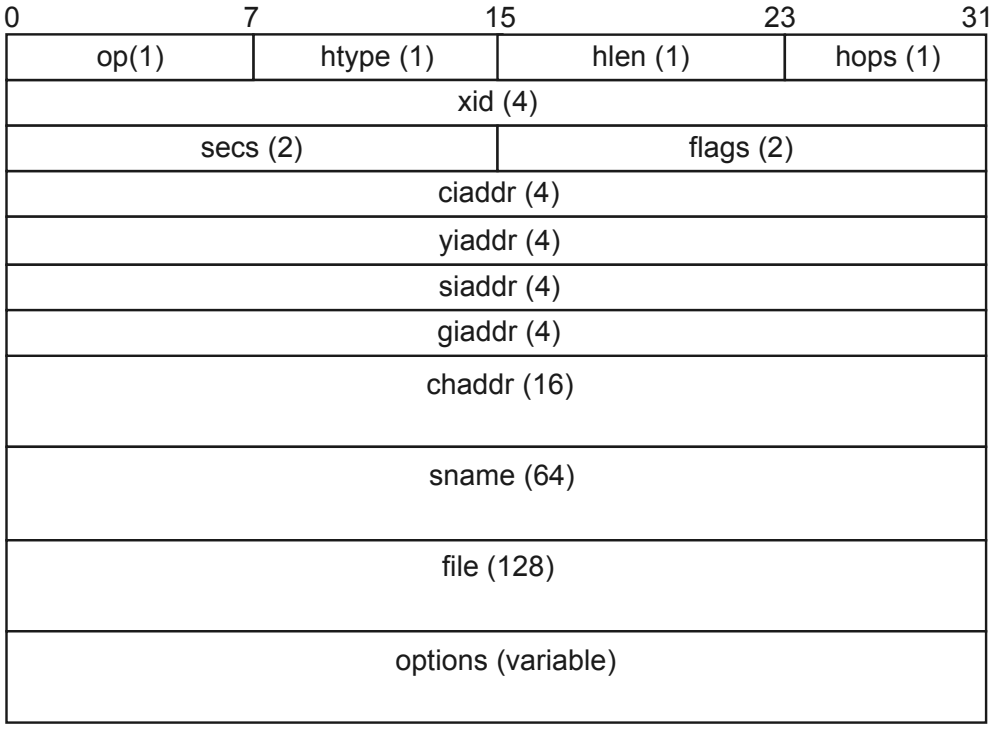
  A DHCP server processes requests of address allocation, address lease extending, and address releasing from a DHCP client or a DHCP relay agent, and allocates IP addresses and other network configuration parameters to the DHCP client.

3.2.2 Introduction to DHCP Messages

DHCP Message Format

Figure 3-2 shows the format of a DHCP message.

Figure 3-2 Format of a DHCP message



In **Figure 3-2**, numbers in the round brackets indicate the field length, expressed in bytes.

Copyright © Huawei Technologies Co., Ltd.

Table 3-1 Description of each field in a DHCP message

Field	Length	Description
op(op code)	1 byte	Indicates the message type. The options are as follows: ● 1: DHCP Request message ● 2: DHCP Reply message
htype (hardware type)	1 byte	Indicates the hardware address type. For Ethernet, the value of this field is 1.
hlen (hardware length)	1 byte	Indicates the length of a hardware address, expressed in bytes. For Ethernet, the value of this field is 6.
hops	1 byte	Indicates the number of DHCP relay agents that a DHCP Request message passes through. This field is set to 0 by a DHCP client. The value increases by 1 each time a DHCP Request message passes through a DHCP relay agent. This field limits the number of DHCP relay agents that a DHCP message can pass through. **NOTE** A maximum of 16 DHCP relay agents are allowed between a server and a client. That is, the number of hops must be smaller than or equal to 16. Otherwise, DHCP messages are discarded.
xid	4 bytes	Indicates a random number chosen by a DHCP client. It is used by the DHCP client and DHCP server to exchange messages.
secs (seconds)	2 bytes	Indicates the period elapsed since a DHCP client began to request an IP address, expressed in seconds.
flags	2 bytes	Indicates the Flags field. Only the leftmost bit of the Flags field is valid and other bits are set to 0. The leftmost bit determines whether the DHCP server unicasts or broadcasts a DHCP Reply message. The options are as follows: ● 0: The DHCP server unicasts a DHCP Reply message. ● 1: The DHCP server broadcasts a DHCP Reply message.
ciaddr (client ip address)	4 bytes	Indicates the IP address of a client. The IP address can be an existing IP address of a DHCP client or an IP address assigned by a DHCP server to a DHCP client. During initialization, the client has no IP address and the value of this field is 0.0.0.0. **NOTE** The IP address 0.0.0.0 is used only for temporary communication during system startup in DHCP mode. It is an invalid address.
yiaddr (your client ip address)	4 bytes	Indicates the DHCP client IP address assigned by the DHCP server. The DHCP server fills this field into a DHCP Reply message.
siaddr (server ip address)	4 bytes	Server IP address from which a DHCP client obtains the startup configuration file.

Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
giaddr (gateway ip address)	4 bytes	Indicates the IP address of the first DHCP relay agent. If the DHCP server and client are located on different network segments, the first DHCP relay agent fills its IP address into this field of the DHCP Request message sent by the client and forwards the message to the DHCP server. The DHCP server determines the network segment where the client resides based on this field, and assigns an IP address on this network segment from an address pool. The DHCP server also returns a DHCP Reply message to the first DHCP relay agent. The DHCP relay agent then forwards the DHCP Reply message to the client. **NOTE** If the DHCP Request message passes through multiple DHCP Relay agents before reaching the DHCP server, the value of this field is the IP address of the first DHCP relay agent and remains unchanged. However, the value of the Hops field increases by 1 each time a DHCP Request message passes through a DHCP relay agent.
chaddr (client hardware address)	16 bytes	Indicates the client MAC address. This field must be consistent with the hardware type and hardware length fields. When sending a DHCP Request message, the client fills its hardware address into this field. For Ethernet, a 6-byte Ethernet MAC address must be filled in this field when the hardware type and hardware length fields are set to 1 and 6 respectively.
sname (server host name)	64 bytes	Indicates the name of the server from which a client obtains configuration parameters. This field is optional and is filled in by the DHCP server. The field must be filled in with a character string that ends with 0.
file (file name)	128 bytes	Indicates the Bootfile name specified by the DHCP server for a DHCP client. This field is filled in by the DHCP server and is delivered to the client when the IP address is assigned to the client. This field is optional. The field must be filled in with a character string that ends with 0.
options	Variable	Indicates the DHCP Options field. It must be of at least 312 bytes. This field contains the DHCP message type and configuration parameters assigned by a server to a client, including the gateway IP address, DNS server IP address, and IP address lease. For details about the Options field, see **3.2.3 DHCP Options**.

DHCP Message Types

DHCP messages are classified into eight types. A DHCP server and a DHCP client communicate by exchanging DHCP messages.

Copyright © Huawei Technologies Co., Ltd.

Table 3-2 DHCP message types

Message Name	Description
DHCP DISCOVER	A DHCP Discover message is broadcast by a DHCP client to locate a DHCP server when the client attempts to connect to a network for the first time.
DHCP OFFER	A DHCP Offer message is sent by a DHCP server to respond to a DHCP Discover message. A DHCP Offer message carries various configuration information.
DHCP REQUEST	A DHCP Request message is sent in the following conditions: ● After a DHCP client is initialized, it broadcasts a DHCP Request message to respond to the DHCP Offer message sent by a DHCP server. ● After a DHCP client restarts, it broadcasts a DHCP Request message to confirm the configuration including the assigned IP address. ● After a DHCP client obtains an IP address, it unicasts or broadcasts a DHCP Request message to update the IP address lease.
DHCP ACK	A DHCP ACK message is sent by a DHCP server to acknowledge the DHCP Request message from a DHCP client. After receiving a DHCP ACK message, the DHCP client obtains the configuration parameters including the IP address.
DHCP NAK	A DHCP NAK message is sent by a DHCP server to reject the DHCP Request message from a DHCP client. For example, after a DHCP server receives a DHCP Request message, it cannot find matching lease records. Then the DHCP server sends a DHCP NAK message, notifying that no IP address is available for the DHCP client.
DHCP DECLINE	A DHCP Decline message is sent by a DHCP client to notify the DHCP server that the assigned IP address conflicts with another IP address. Then the DHCP client applies to the DHCP server for another IP address.
DHCP RELEASE	A DHCP Release message is sent by a DHCP client to release its IP address. After receiving a DHCP Release message, the DHCP server can assign this IP address to another DHCP client.
DHCP INFORM	A DHCP Inform message is sent by a DHCP client to obtain other network configuration parameters such as the gateway address and DNS server address after the DHCP client has obtained an IP address.

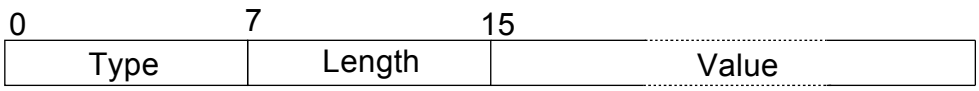
3.2.3 DHCP Options

Options Field in a DHCP Packet

The Options field in a DHCP packet carries control information and parameters that are not defined in common protocols. When a DHCP client requests an IP address from the DHCP server configured with the Options field, the server returns a DHCP Reply packet containing the Options field. **Figure 3-3** shows the format of the Options field.

Copyright © Huawei Technologies Co., Ltd.

Figure 3-3 Format of the Options field



The Options field consists of Type, Length, and Value. The following table provides the details.

Table 3-3 Description of the Options field

Field	Length	Description
Type	1 byte	Indicates the type of the message content.
Length	1 byte	Indicates the length of the message content.
Value	Depending on the setting of the Length field	Indicates the message content.

The value of the Options field ranges from 1 to 255. **Table 3-4** lists common DHCP options.

Table 3-4 Description of the Options field in DHCP packets

Options No.	Function
1	Specifies the subnet mask.
3	Specifies the gateway address.
6	Specifies the DNS server IP address.
12	Specifies the hostname.
15	Specifies the domain name.
33	Specifies a group of classful static routes. This option contains a group of classful static routes. When a DHCP client receives DHCP packets with this option, it adds the classful static routes contained in the option to its routing table. In classful routes, masks of destination addresses are natural masks and masks cannot be used to divide subnets. If Option 121 exists, this option is ignored.
44	Specifies the NetBIOS name.
46	Specifies the NetBIOS object type.
50	Specifies the requested IP address.
51	Specifies the IP address lease.

Copyright © Huawei Technologies Co., Ltd.

Options No.	Function
52	Specifies the additional option.
53	Specifies the DHCP packet type.
54	Specifies the server identifier.
55	Specifies the parameter request list. It is used by a DHCP client to request specified configuration parameters.
58	Specifies the lease renewal time (T1), which is 50% of the lease time.
59	Specifies the lease renewal time (T2), which is 87.5% of the lease time.
60	Specifies Class Id.
61	Specifies Client Id.
66	Specifies the TFTP server name allocated to DHCP clients.
67	Specifies the Bootfile name allocated to DHCP clients.
77	Specifies the user type.
121	Specifies a group of classless routes. This option contains a group of classless static routes. After a DHCP client receives DHCP packets with this option, it adds the classless static routes contained in the option to its routing table. Classless routes are routes of which masks of destination addresses can be any values and masks can be used to divide subnets.
148	Commander IP address.
149	The FTPS and SFTP server address.
150	Specifies the TFTP server address allocated to DHCP clients.

The objects of this field vary with the functions of the Options field. For example, Option 77 is used on a DHCP client to identify user types of the DHCP client. The DHCP server selects an address pool to allocate an IP address and configuration parameters to the DHCP client based on the User Class in the Option field. Option 77 is manually configured only on the DHCP client but not on the server.

NOTE

When the device functions as the DHCP client, the client can identify the Option121 field describing static routes in the DHCP packet sent by the DHCP server.

For more information about common DHCP options, see RFC 2132.

Customized DHCP Options

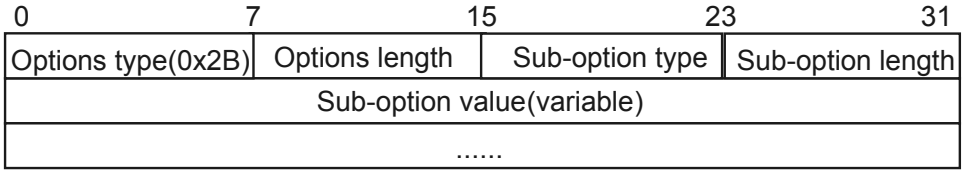
Some options are not defined in RFC 2132. Customized options Option 43 and Option 82 are described as follows:

Copyright © Huawei Technologies Co., Ltd.

- Option 43

  Option 43 is called vendor-specific information option. **Figure 3-4** shows the format of Option 43.

  Figure 3-4 Format of Option 43

  DHCP servers and DHCP clients use Option 43 to exchange vendor-specific information. When a DHCP server receives a DHCP Request packet with parameter 43 encapsulated in Option 55, it encapsulates Option 43 in a DHCP Reply packet and sends it to the DHCP client.

  To implement extensibility and allocate more configuration parameters to DHCP clients, Option 43 supports suboptions, as shown in **Figure 3-4**. Suboptions are described as follows:

  - Sub-option type: The value 0x01 indicates the ACS parameter, the value 0x02 indicates the SP ID, and the value 0x80 indicates the PXE server address.
  - Sub-option length
  - Sub-option value

  If a device functions as a DHCP client, it can obtain the following information using Option 43:

  - Auto-configuration server (ACS) parameters, including the URL, user name, and password
  - SP ID that the Customer Premises Equipment (CPE) notifies the ACS so that the ACS selects configuration parameters from the specified SP
  - Preboot execution environment (PXE) server address, which is used by a DHCP client to obtain the Bootfile or control information from the PXE server

- Option 82

  The Option 82 field is called the DHCP relay agent information field. It records the location of a DHCP client. A DHCP relay agent or a device enabled with DHCP snooping appends the Option 82 field to a DHCP Request message sent from a DHCP client, and then forwards the DHCP Request message to a DHCP server.

  You can use the Option 82 field to locate a DHCP client and implement control security and accounting of the DHCP client. The DHCP server that supports the Option 82 field can determine allocation of IP addresses and other parameters according to the information in the Option 82 field. IP addresses can be assigned flexibly.

  The Option 82 field contains a maximum of 255 suboptions. If the Option 82 field is defined, at least one suboption must be defined. Currently, the device supports only two suboptions: sub-option 1 (circuit ID) and suboption 2 (remote ID).

The content of the Option 82 field is not defined uniformly, and various vendors fill in the Option 82 field as required.

The device supports the following formats for the Option 82 field:

Copyright © Huawei Technologies Co., Ltd.

- Default: It is the default format of the Option 82 field.
- Common: The Option 82 field in common format uses a character string and is for specific markets.
- Extend: It is compatible with formats of Option 82 fields on non-Huawei switches. The Option 82 field in extend format can use binary notation.
- User-defined: This format is used if the format of the Option 82 field is not defined.

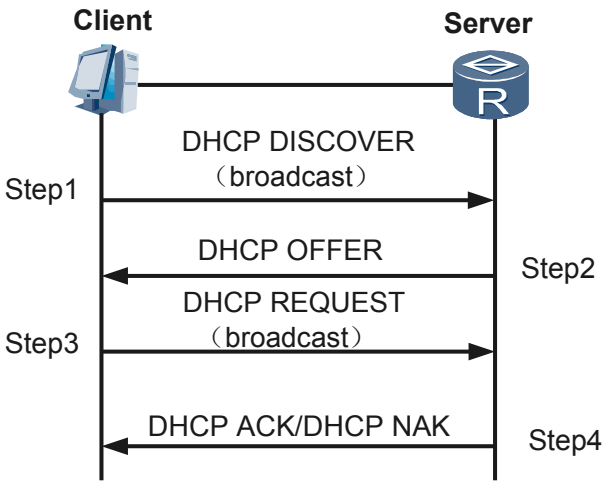
3.2.4 DHCP Principles

Modes for Interaction Between the DHCP Client and Server

To obtain a valid dynamic IP address, a DHCP client exchanges different messages with the server at different stages. Generally, the DHCP client and server interact in the following modes.

- The DHCP client dynamically obtains an IP address.

Figure 3-5 Procedure for a DHCP client to dynamically obtain an IP address



As shown in **Figure 3-5**, when a DHCP client accesses the network for the first time, the DHCP client sets up a connection with a DHCP server through the following four stages.

- Discovery stage: The DHCP client searches for the DHCP server.

  In this stage, the DHCP client sends a DHCP Discover message to search for the DHCP server. The DHCP server address is unknown to the client, so the DHCP client broadcasts the DHCP Discover message. All the DHCP servers send Reply messages after they receive the Discover message. In this way, the DHCP client knows locations of the DHCP servers on the network.

- Offer stage: The DHCP server offers an IP address to the DHCP client.

  The DHCP server receives the DHCP Discover message, selects an IP address from the address pool, and sends a DHCP Offer message to the DHCP client. The Offer message carries information such as the IP address, lease of the IP address, gateway address, and DNS server address.

- Request stage: The DHCP client selects an IP address.

Copyright © Huawei Technologies Co., Ltd.

If multiple DHCP servers send DHCP Offer messages to the DHCP client, the client receives the first DHCP Offer message. Then the client broadcasts a DHCP Request message including the information about the DHCP server address (Option 54 field).

The client broadcasts a DHCP Request message to notify all the DHCP servers that the client uses the IP address provided by the DHCP server in the Option 54 field and that all the other servers can use the assigned IP addresses.

- Acknowledgment stage: The DHCP server acknowledges the IP address that is offered.

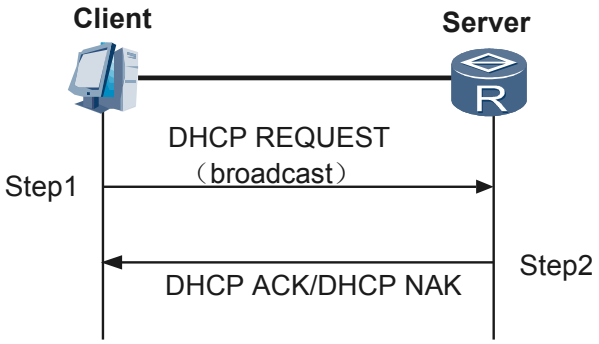
  When the DHCP server receives the DHCP Request message from the DHCP client, the server searches the lease record based on the MAC address in the Request message. If there is the IP address record, the server sends a DHCP ACK message to the client, carrying the IP address and other configurations. After receiving the DHCP ACK message, the DHCP client broadcasts gratuitous ARP packets to detect whether any host is using the IP address assigned by the DHCP server. If no response is received within the specified time, the DHCP client uses the IP address.

  If there is no IP address record or the server cannot assign IP addresses, the server sends a DHCP NAK message to notify the DHCP client that the server cannot assign IP addresses. The DHCP client needs to send a new DHCP Discover message to request a new IP address.

  After obtaining the IP address, the DHCP client checks the status of the gateway in use before the client goes online. If the gateway address is incorrect or the gateway device fails, the DHCP client requests a new IP address using the four modes for interaction.

- The DHCP client uses the assigned IP address.

Figure 3-6 Procedure for the DHCP client to use the assigned IP address



As shown in **Figure 3-6**, when the DHCP client accesses a network for the second time, it set ups a connection with the DHCP server in the following procedure.

- The client accesses a network for the second time with the IP address that does not expire. The client does not need to send a DHCP Discover message again. It directly sends a DHCP Request message carrying the IP address assigned in the first time, namely, the Option 50 field in the message.
- After receiving the DHCP Request message, if the requested IP address is not assigned to another DHCP client, the DHCP server sends a DHCP ACK message to instruct the DHCP client to use the IP address again.
- If the IP address cannot be assigned to the DHCP client, for example, it has been assigned to another DHCP client, the DHCP server sends a DHCP NAK message to the DHCP

Copyright © Huawei Technologies Co., Ltd.

client. After receiving the DHCP NAK message, the DHCP client sends a DHCP Discover message to request a new IP address.

- The DHCP client renews the IP address lease.

  An expected lease can be contained in the DHCP Request message sent to the server for an IP address. The server compares the expected lease with the lease in the address pool and assigns a shorter lease to the client.

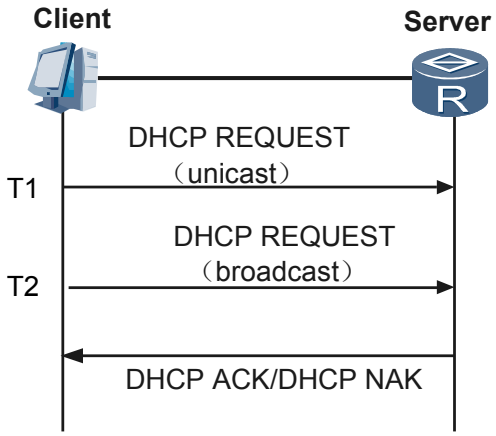
  The IP address dynamically assigned to the DHCP client usually has a validity period. The DHCP server withdraws the IP address after the validity period expires. To keep using the IP address, the DHCP client needs to renew the IP address lease.

  When obtaining an IP address, the DHCP client enters the binding state. The client is configured with three timers to control lease renewal, rebinding, and lease expiration respectively. When assigning an IP address to the DHCP client, the DHCP server also specifies values for the timers. If the server does not specify values for the timers, the client uses the default values. **Table 3-5** lists the default timer values.

Table 3-5 Default values of timers

Timer	Default Value
Lease renewal	50% of the lease
Rebinding	87.5% of the lease
Lease expiration	Overall lease

Figure 3-7 Procedure for a DHCP client to renew the IP address lease



As shown in **Figure 3-7**, when the DHCP client renews the IP address lease, it set ups a connection with the DHCP server in the following procedures:

- When 50% of the IP address lease (T1) has passed, the DHCP client unicasts a DHCP Request message to the DHCP server to renew the lease. If the client receives a DHCP ACK message, the address lease is successfully renewed. If the client receives a DHCP NAK message, it sends a request again.
- When 87.5% of the IP address lease (T2) has passed and the client has not received the Reply message, the DHCP client automatically sends a broadcast message to the DHCP

Copyright © Huawei Technologies Co., Ltd.

server to renew the IP address lease. If the client receives a DHCP ACK message, the address lease is successfully renewed. If the client receives a DHCP NAK message, it sends a request again.

- If the client has not received a Reply message from the server when the IP address lease expires, the client must stop using the current IP address and send a DHCP Discover message to request a new IP address.

- The DHCP client releases an IP address.

  When the DHCP client does not use the assigned IP address, it sends a DHCP Release message to notify the DHCP server of releasing the IP address. The DHCP server retains the DHCP client configurations so that the configurations can be used when the client requests an address again.

3.2.5 DHCP Relay Principles

The DHCP relay function enables message exchanges between a DHCP server and a client on different network segments. When the DHCP client and server are on different network segments, the DHCP relay agent transparently transmits DHCP messages to the destination DHCP server. In this way, DHCP clients on different network segments can communicate with one DHCP server.

Figure 3-8 shows how a DHCP client uses the DHCP relay agent to apply for an IP address for the first time.

Figure 3-8 Working process of a DHCP relay agent

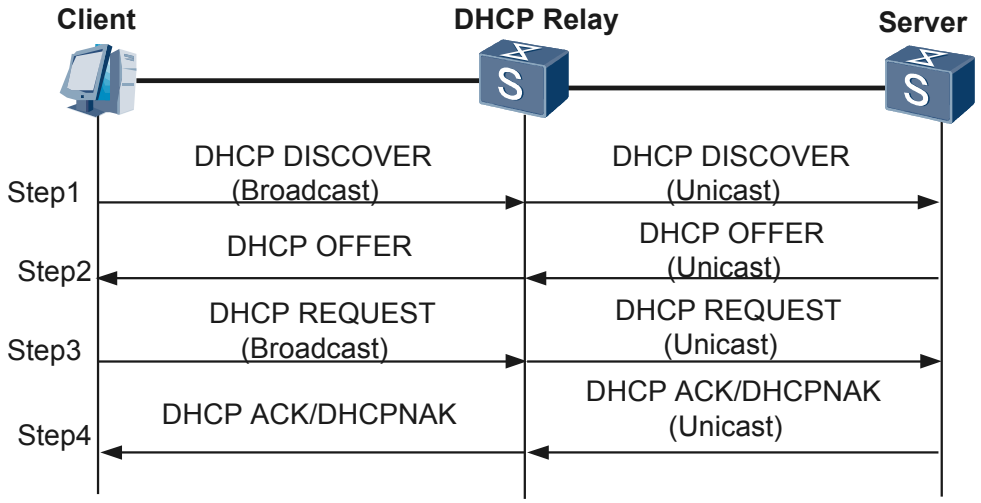


Figure 3-8 shows the working process of a DHCP relay agent. The DHCP client sends a Request message to the DHCP server. When receiving the message, the DHCP relay agent processes and unicasts the message to the specified DHCP server on the other network segment. The DHCP server sends requested configurations to the client through the DHCP relay agent based on information in the Request message.

1. After receiving a DHCP Discover message or a Request message, the DHCP relay agent performs the following operations:

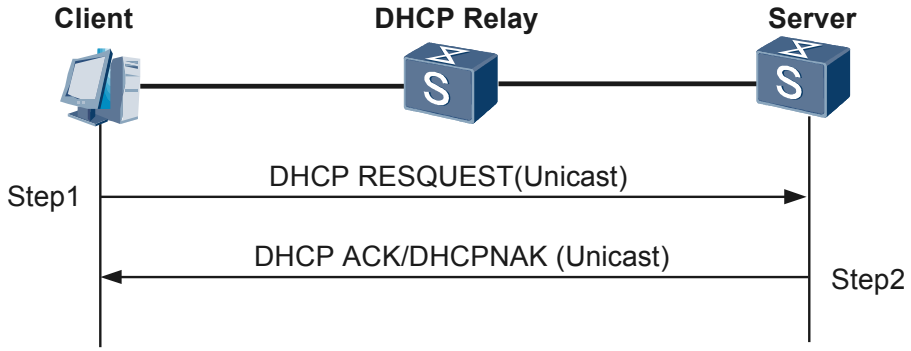
Copyright © Huawei Technologies Co., Ltd.

- Discards DHCP Request messages whose number of hops is larger than the hop limit to prevent loops. Or, increases the value of the hop by 1, indicating that the message passes through a DHCP relay agent.
- Checks the Relay Agent IP Address field. If the value is 0, set the value of the Relay Agent IP Address field to the IP address of the interface which receives the Request message. Selects one IP address if the interface has multiple IP addresses. All the Request messages received by the interface later use this IP address to fill the Relay Agent IP Address field. If the value is not 0, do not change the value.
- Sets the TTL value of the Request message to the default TTL value of the DHCP relay agent. You can change the value of the hops field to prevent loops and limit hops.
- Changes the destination IP address of the DHCP Request message to the IP address of the DHCP server or the IP address of the next DHCP relay agent. In this way, the DHCP Request message can be forwarded to the DHCP server or the next DHCP relay agent.

2. The DHCP server assigns IP addresses to the client based on the Relay Agent IP Address field and sends the DHCP Reply message to the DHCP relay agent specified in the Relay Agent IP Address field. After receiving the DHCP Reply message, the DHCP relay agent performs the following operations:
   - The DHCP relay agent assumes that all the Reply messages are sent to the directly-connected DHCP clients. The Relay Agent IP Address field identifies the interface directly connected to the client. If the value of the Relay Agent IP Address field is not the IP address of a local interface, the DHCP relay agent discards the Reply message.
   - The DHCP relay agent checks the broadcast flag bit of the message. If the broadcast flag bit is 1, the DHCP relay agent broadcasts the DHCP Reply message to the DHCP client; otherwise, the DHCP relay agent unicasts the DHCP Reply message to the DHCP client. The destination IP address is the value in the Your (Client) IP Address field, and the MAC address is the value in the Client Hardware Address field.

Figure 3-9 shows how a DHCP client extends the IP address lease through the DHCP relay agent.

Figure 3-9 Extending the IP address lease through the DHCP relay agent



1. After accessing the network for the first time, the DHCP client only needs to unicast a DHCP Request message to the DHCP server that assigned its currently-used IP address.
2. The DHCP server then directly unicasts a DHCP ACK message or a DHCP NAK message to the client.

Copyright © Huawei Technologies Co., Ltd.

DHCP Releasing

The DHCP relay agent, instead of the client, can send a Release message to the DHCP server to release the IP addresses that assigned to the DHCP clients. You can configure a command on the DHCP relay agent to release the IP addresses that the DHCP server assigns to the DHCP client.

3.2.6 IP Address Assignment and Renewal

IP Address Assignment Sequence

The DHCP server assigns IP addresses to a client in the following sequence:

- IP address that is in the database of the DHCP server and is statically bound to the MAC address of the client
- IP address that has been assigned to the client before, that is, IP address in the Requested IP Addr Option of the DHCP Discover message sent by the client
- IP address that is first found when the DHCP server searches the DHCP address pool for available IP addresses
- If the DHCP address pool has no available IP address, the DHCP server searches the expired IP addresses and conflicting IP addresses, and then assigns a valid IP address to the client. If all the IP addresses are in use, an error is reported.

Method of Preventing Repeated IP Address Assignment

Before assigning an IP address to a client, the DHCP server needs to ping the IP address to avoid address conflicts.

By using the ping command, you can check whether a response to the ping packet is received within the specified period. If no response to the ping packet is received, the DHCP server keeps sending ping packets to the IP address to be assigned until the number of the sent ping packets reaches the maximum value. If there is still no response, this IP address is not in use, and the DHCP server assigns the IP address to a client. (This is implemented based on RFC 2132.)

IP Address Reservation

DHCP supports IP address reservation for clients. The reserved IP addresses can be those in the address pool or not. If an address in the address pool is reserved, it is no longer assignable. Addresses are usually reserved for DNS servers.

Method of IP Address Releasing and Lease Renewal on the PCs

The PCs (DHCP clients) must release the original IP addresses before obtaining new IP addresses.

- Releasing the original IP address

  Commands for renewing the lease of an IP address vary in different operating systems. You can use either of the following methods to renew the lease of an IP address:

  - Run the **ipconfig**/**release** command in the Window Vista/Windows XP/Windows2000/ DOS environment of the user PC to release the IP address of the PC.

Copyright © Huawei Technologies Co., Ltd.

- Run the **winipcfg**/**release** command in the MS-DOS interface of Windows 98 to release the IP address of the PC.

The user PC needs to send a DHCP Release message to the DHCP server.

- Renewing the IP address lease or applying for a new IP address

  The same command is used to apply for a new IP address and renew the IP address in the same operating system. Before applying for a new IP address, the PCs (DHCP clients) must release the original IP addresses. If you want to renew the IP address lease, you do not have to release the IP address.

  Different commands are used in different operating systems. You can use either of the following methods to apply for a new IP address:

  - Run the **ipconfig**/**renew** command in the Windows Vista/Windows XP/Windows2000/ DOS environment of the user PC to apply for a new IP address.
  - Run the **winipcfg**/**renew** command in the MS-DOS interface of Windows 98 to apply for a new IP address.

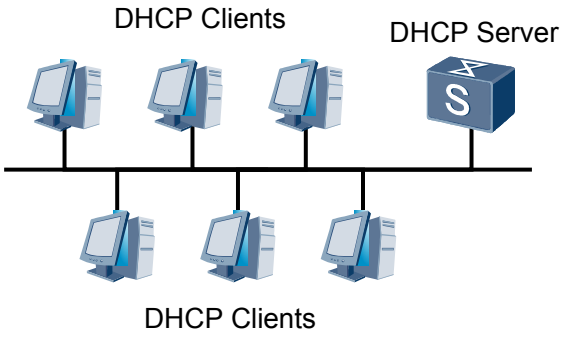
  The user PC needs to send a DHCP Discover message to the DHCP server.

3.3 Application

3.3.1 DHCP Server Application

As it is shown in **Figure 3-10**, a DHCP server and multiple DHCP clients (such as PCs and portable computers) are deployed.

Figure 3-10 Typical networking of the DHCP server



Generally, the DHCP server is used to assign IP addresses in the following scenarios:

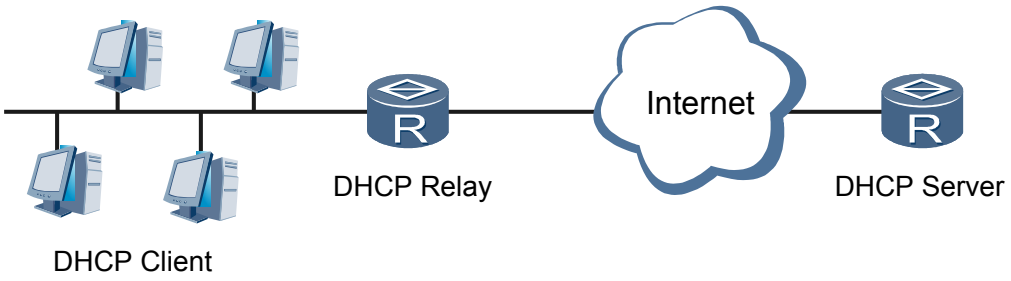
- On a large network, manual configurations take a long time and bring difficulties to centralized management over the entire network.
- Hosts on the network are more than available IP addresses. Thus, not every host has a fixed IP address. Many hosts need to dynamically obtain IP addresses through the DHCP server. In addition, network administrators hope that there is a limit to the number of users of on-line at the same time.
- Only a few hosts on the network require fixed IP addresses.

Copyright © Huawei Technologies Co., Ltd.

3.3.2 DHCP Relay Application

Figure 3-11 shows typical networking of DHCP relay.

Figure 3-11 Typical networking of DHCP relay



The earlier DHCP protocol applies to only the scenario that the DHCP client and DHCP server are on the same network segment. To dynamically assign IP addresses to hosts on network segments, the network administrator needs to configure a DHCP server on each network segment, which increases costs.

The DHCP relay function is introduced to solve this problem. A DHCP client can apply to the DHCP server on another network segment to obtain a valid IP address. In this manner, DHCP clients on multiple network segments can share one DHCP server. This reduces costs and facilitates centralized management.

Generally, a host, a Layer 3 switch, or a router can function as a DHCP relay agent only after it is enabled with DHCP relay.

3.4 References

The following table lists the references of this document.

Document	Description	Remarks
RFC1533	DHCP Options and BOOTP Vendor Extensions	-
RFC1534	Interoperation Between DHCP and BOOTP	-
RFC2131	Dynamic Host Configuration Protocol	-
RFC2132	DHCP Options and BOOTP Vendor Extensions	-
RFC3046	DHCP Relay Agent Information Option	-

Copyright © Huawei Technologies Co., Ltd.

4 DHCPv6

About This Chapter

4.1 Introduction
4.2 Principles
4.3 References

Copyright © Huawei Technologies Co., Ltd.

4.1 Introduction

Definition

Dynamic Host Configuration Protocol for IPv6 (DHCPv6) is designed to assign IPv6 addresses, prefixes, and other network configuration parameters to hosts.

Purpose

The IPv6 protocol provides huge address space formed by 128-bit IPv6 addresses that require proper and efficient assignment and management policies. IPv6 stateless address autoconfiguration defined in RFC2462 is widely used. Hosts configured with the stateless address autoconfiguration function automatically configure IPv6 addresses based on prefixes carried in Route Advertisement (RA) packets sent from a neighboring router.

When stateless address autoconfiguration is used, routers do not record IPv6 addresses of hosts. Therefore, stateless address autoconfiguration has poor manageability. In addition, hosts configured with stateless address autoconfiguration cannot obtain other configuration parameters such as the DNS server address. ISPs do not provide instructions for automatic allocation of IPv6 prefixes for routers. Therefore, users need to manually configure IPv6 addresses for routing and switching devices during IPv6 network deployment.

DHCPv6 solves this problem. DHCPv6 is a stateful protocol for configuring IPv6 addresses automatically. During stateful address configuration, a DHCPv6 server assigns a complete IPv6 address to a host and provides other configuration parameters, such as the DNS server address. A DHCPv6 relay agent may be used to relay DHCPv6 packets. The DHCPv6 server binds the IPv6 address to a client. This improves network manageability.

Compared with manual address configuration and IPv6 stateless address autoconfiguration that uses network prefixes in RA packets, DHCPv6 has the following advantages:

- Controls IPv6 address assignment better. A DHCPv6 device can record addresses assigned to hosts and assign requested addresses. This function facilitates network management.
- Assigns IPv6 address prefixes to network devices. This function facilitates automatic configuration and hierarchical network management.
- Provides other network configuration parameters such as the DNS server address.

4.2 Principles

4.2.1 DHCPv6 Overview

DHCPv6 runs between a client and a server. Similar to DHCP for IPv4, DHCPv6 clients and DHCPv6 servers exchange DHCPv6 packets using the User Datagram Protocol (UDP). In IPv6, packets cannot be broadcast; therefore, DHCPv6 uses multicast packets. In this case, DHCPv6 clients do not need to be configured with IPv6 addresses of DHCPv6 servers.

Copyright © Huawei Technologies Co., Ltd.

IPv6 Address Allocation Methods

The IPv6 protocol provides huge address space formed by 128-bit IPv6 addresses that require proper and efficient assignment and management policies.

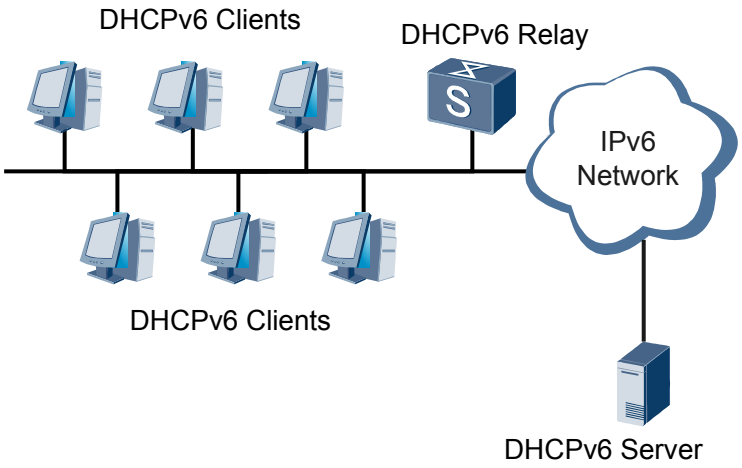
Currently, the following methods are available to allocate IPv6 addresses:

- Manual configuration: You can manually configure IPv6 addresses, prefixes, and other network configuration parameter, such as addresses of the Domain Name System (DNS), Network Information Service (NIS), and Simple Network Time Protocol (SNTP) servers.
- Stateless address autoconfiguration: Hosts generate a link-local address based on the interface ID and automatically configure IPv6 addresses based on prefixes carried in RA packets.
- stateful autoconfiguration,that is DHCPv6,DCHPv6 allocation has the following three methods:
  - DHCPv6 stateful autoconfiguration: DHCPv6 servers automatically configure IPv6 addresses, prefixes, and other network configuration parameters, such as addresses of the DNS, NIS, and SNTP servers.
  - DHCPv6 stateless autoconfiguration: IPv6 addresses are generated based on RA packets. A DHCPv6 server does not provide IPv6 addresses but provides other configuration parameters about the DNS, NIS, and SNTP servers.
  - DHCPv6 PD prefix autoconfiguration: A downstream router requests IPv6 prefixes from an upstream router. The upstream router assigns requested prefixes to the downstream router. You do not need to manually configure IPv6 prefixes for user-side links of the downstream router. The downstream router divides the obtained prefix (the length of the obtained prefix is smaller than 64 bits) into 64-bit prefix of subnet segments and sends an RA packet on the link that IPv6 hosts directly connect to. This enables hosts to automatically configure addresses, completing IPv6 network deployment.

DHCPv6 Architecture

Figure 4-1 shows the DHCPv6 architecture.

Figure 4-1 DHCPv6 architecture



DHCPv6 involves the following roles:

Copyright © Huawei Technologies Co., Ltd.

- DHCPv6 client

  A DHCPv6 client applies to a DHCPv6 server for IPv6 addresses, prefixes, and network configuration parameters to complete its address configuration.

- DHCPv6 relay

  A DHCPv6 relay agent relays DHCPv6 packets between a DHCPv6 client and a DHCPv6 server to help the DHCPv6 client complete its address configuration. Generally, a DHCPv6 client communicates with a DHCPv6 server through the link-local multicast address to obtain IPv6 addresses, prefixes, and other network configuration parameters. If a DHCPv6 server and a DHCPv6 client are on different links, a DHCPv6 relay agent is required to forward DHCPv6 packets. In this case, you do not need to deploy a DHCPv6 server on each link, which saves costs and facilitates centralized management.

  A DHCPv6 relay agent is optional. If a DHCPv6 client and a DHCPv6 server are on the same link or a DHCPv6 client communicates with a DHCPv6 server in unicast mode to complete address allocation or information configuration, you do not need to deploy a DHCPv6 relay agent. A DHCPv6 relay agent is required only when a DHCPv6 client and a DHCPv6 server are located on different links or a DHCPv6 client cannot communicate with a DHCPv6 server in unicast mode.

- DHCPv6 server

  A DHCPv6 server processes requests of address allocation, address lease extension, and address release from a DHCPv6 client or a DHCPv6 relay agent, and assigns IPv6 addresses and other network configuration parameters to the DHCPv6 client.

Basic DHCPv6 Concepts

1. Multicast address
   - In DHCPv6, a DHCPv6 client does not need to be configured with the IPv6 address of a DHCPv6 server. Instead, the DHCPv6 client locates DHCPv6 servers by sending Solicit packets with multicast addresses as destination addresses.
   - In DHCPv4, a DHCP client locates DHCP servers by broadcasting DHCP packets. To prevent broadcast storms, IPv6 does not use broadcast packets. Instead, IPv6 uses multicast packets. DHCPv6 uses the following two multicast addresses:
     - FF02::1:2 (All DHCP Relay Agents and Servers): indicates the multicast address of all the DHCPv6 servers and DHCPv6 relay agents. The address is a link-local multicast address and is used for communication between a DHCPv6 client and its neighboring servers or between a DHCPv6 client and DHCPv6 relay agents. All DHCPv6 servers and relay agents are members of this multicast group.
     - FF05::1:3 (All DHCP Servers): indicates the multicast address of all the DHCPv6 servers. The address is a site-local address and is used for communication between DHCPv6 relay agents and DHCPv6 servers within a site. All DHCPv6 servers within a site are members of this multicast group.
2. UDP port number
   - DHCPv6 packets are transmitted through UDPv6.
   - DHCPv6 clients only process DHCPv6 packets with UDP port number 546.
   - DHCPv6 servers and relay agents only process DHCPv6 packets with UDP port number 547.
3. DHCP Unique Identifier (DUID)

Copyright © Huawei Technologies Co., Ltd.

- A DUID identifies a DHCPv6 device. Each DHCPv6 server or client has a unique DUID. DHCPv6 servers use DUIDs to identify DHCPv6 clients and DHCPv6 clients use DUIDs to identify DHCPv6 servers.
- The DUIDs of a DHCPv6 client and a DHCPv6 server are carried in the Client Identifier option and the Server Identifier option respectively. The Client Identifier option and the Server Identifier option have the same format and are distinguished by the option-code field value.

4. Identity association (IA)
   - An IA enables a DHCPv6 server and a DHCPv6 client to identify, group, and manage IPv6 addresses. Each IA consists of an identity association identifier (IAID) and associated configuration information.
   - A DHCPv6 client must associate at least one IA with each of its network interfaces for which the DHCPv6 client requests IPv6 addresses from a DHCP server. The DHCPv6 client uses IAs associated with network interfaces to obtain configuration information from a DHCPv6 server. Each IA must be associated with at least one interface.
   - The IAID identifies an IA, and IAIDs on the same DHCPv6 client must be unique. The IAID is not lost or changed because of factors such as DHCPv6 client reboot.
   - The configuration information in an IA consists of one or more IPv6 addresses along with the lifetimes T1 and T2. Each address in an IA has a preferred lifetime and a valid lifetime.
   - An interface must be associated with at least one IA; an IA can contain information about one or more addresses.

4.2.2 DHCPv6 Packets

DHCPv6 Packet Format

Figure 4-2 shows the DHCPv6 packet format.

Figure 4-2 DHCPv6 packet format

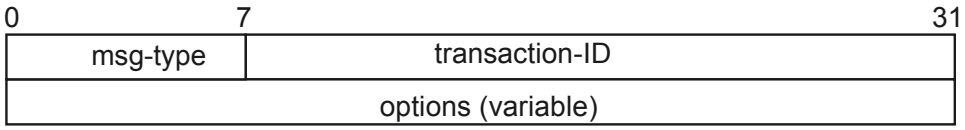


Table 4-1 Description of each field in a DHCPv6 packet

Field	Length	Description
msg-type	1 byte	Indicates the packet type. The value ranges from 1 to 13. For details, see the **DHCPv6 Packet Type**.

Copyright © Huawei Technologies Co., Ltd.

Field	Length	Description
transaction-ID	3 bytes	Identifies packet transaction between DHCPv6 clients and servers. For example, a DHCPv6 client initiates a Solicit/Advertise transaction or a Request/Reply transaction. Their transaction IDs are different. Transaction IDs have the following characteristics: ● The transaction ID is randomly generated by a DHCPv6 client. ● Transaction IDs of Request and Reply packets must be the same. ● The transaction ID of a packet initiated by a DHCPv6 server is 0.
Options	Variable	Indicates the option field in a DHCPv6 packet. The option field contains configurations that the DHCPv6 server assigns to IPv6 hosts. The configurations include the IPv6 address of the DNS server.

DHCPv6 Packet Type

DHCPv6 defines 13 types of packets. A DHCPv6 server and a DHCPv6 client communicate by exchanging these types of packets. The following table lists DHCPv6 packets and their corresponding DHCPv4 packets and describes the DHCPv6 packets.

DHCP Packet Type	DHCPv6 Packet	DHCPv4 Packet	Description
1	SOLICIT	DHCP DISCOVER	A DHCPv6 client sends a Solicit packet to locate DHCPv6 servers.
2	ADVERTISE	DHCP OFFER	A DHCPv6 server sends an Advertise packet in response to a Solicit packet to declare that it can provide DHCPv6 services.
3	REQUEST	DHCP REQUEST	A DHCPv6 client sends a Request packet to request IPv6 addresses and other configuration parameters from a DHCPv6 server.
4	CONFIRM	-	A DHCPv6 client sends a Confirm packet to any available DHCPv6 server to check whether the obtained IPv6 address applies to the link that the DHCPv6 client is connected to.
5	RENEW	DHCP REQUEST	A DHCPv6 client sends a Renew packet to the DHCPv6 server that provides the IPv6 addresses and other configuration parameters to extend the lifetime of the addresses and to update configuration parameters.

Copyright © Huawei Technologies Co., Ltd.

DHCP Packet Type	DHCPv6 Packet	DHCPv4 Packet	Description
6	REBIND	DHCP REQUEST	A DHCPv6 client sends a Rebind packet to any available DHCPv6 server to extend the lifetime of the assigned IPv6 address and to update configuration parameters when the client does not receive a response to its Renew packet.
7	REPLY	DHCP ACK/ NAK	A DHCPv6 server sends a Reply packet in the following situations: 1. A DHCPv6 server sends a Reply packet containing IPv6 addresses and configuration parameters in response to a Solicit, Request, Renew or Rebind packet received from a DHCPv6 client. 2. A DHCPv6 server sends a Reply packet containing configuration parameters in response to an Information-Request packet. 3. A DHCPv6 server sends a Reply packet in response to a Confirm, Release, or Decline packet received from a DHCPv6 client.
8	RELEASE	DHCP RELEASE	A DHCPv6 client sends a Release packet to the DHCPv6 server that assigns IPv6 addresses to the DHCPv6 client, indicating that the DHCPv6 client will no longer use the obtained addresses.
9	DECLINE	DHCP DECLINE	A DHCPv6 client sends a Decline packet to a DHCPv6 server, indicating that the IPv6 addresses assigned by the DHCPv6 server are already in use on the link to which the DHCPv6 client is connected.
10	RECONFIGURE	-	A DHCPv6 server sends a Reconfigure packet to a DHCPv6 client, informing the DHCPv6 client that the DHCPv6 server has new addresses or updated configuration parameters.
11	INFORMATION-REQUEST	DHCP INFORM	A DHCPv6 client sends an Information-Request packet to a DHCPv6 server to request configuration parameters except for IPv6 addresses.
12	RELAY-FORW	-	A DHCPv6 relay agent sends a Relay-Forward packet to relay Request packets to DHCPv6 servers.
13	RELAY-REPL	-	A DHCPv6 server sends a Relay-Reply packet to a DHCPv6 relay agent. The Relay-Reply packet carries a packet that the DHCPv6 relay agent needs to deliver to a DHCPv6 client.

Copyright © Huawei Technologies Co., Ltd.

4.2.3 DHCPv6 Working Principles

DHCPv6 autoconfiguration is classified as stateful or stateless.

- DHCPv6 stateful autoconfiguration: A DHCPv6 server automatically configures IPv6 addresses, prefixes, and network configuration parameters of the DNS, NIS, and SNTP servers.
- DHCPv6 stateless autoconfiguration: IPv6 addresses are generated based on the Route Advertisement (RA) packets. A DHCPv6 server provides other configuration parameters such as addresses of the DNS, NIS, and SNTP servers except for IPv6 addresses.

DHCPv6 Stateful Autoconfiguration

The IPv6 node obtains addresses and other configuration parameters (such as the IPv6 address of the DNS server) through stateful DHCPv6 autoconfiguration.

A DHCPv6 server assigns addresses and prefixes to a DHCPv6 client in the following ways:

- DHCPv6 four-message exchange
- DHCPv6 two-message exchange

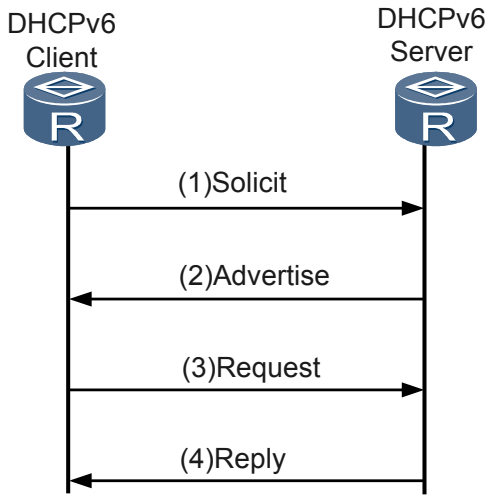
DHCPv6 Four-Message Exchange

Four-message exchange applies to a network where multiple DHCPv6 servers are available. A DHCPv6 client first multicasts a Solicit packet to locate DHCPv6 servers that can provide DHCPv6 services. After receiving Advertise packets from multiple DHCPv6 servers, the DHCPv6 client selects one of the DHCPv6 servers according to priorities of DHCPv6 servers. Then the DHCPv6 client and the selected DHCPv6 server complete address application and allocation by exchanging Request and Reply packets.

If a DHCPv6 server does not have two-message exchange enabled, the DHCPv6 server allocates addresses and configuration parameters through four-message exchange, regardless of whether the Solicit packet contains the Rapid Commit option.

Figure 4-3 shows the process of address allocation using four-message exchange.

Figure 4-3 Process of address allocation using four-message exchange



Copyright © Huawei Technologies Co., Ltd.

The process of address allocation using four-message exchange is as follows:

1. A DHCPv6 client sends a Solicit packet to request a DHCPv6 server to allocate IPv6 addresses and network configuration parameters.
2. If the DHCPv6 server does not support fast address allocation, the DHCPv6 server returns an Advertise packet containing the allocated addresses and network configuration parameters regardless of whether the Solicit packet contains the Rapid Commit option.
3. If receiving Advertise packets from multiple DHCPv6 servers, the DHCPv6 client selects the DHCPv6 server with the highest priority and sends Request multicast packets to all DHCPv6 servers. The Request multicast packets carry the DUID of the selected DHCPv6 server.
4. The DHCPv6 server responds with a Reply packet that contains the addresses and network configuration parameters allocated to the client.

DHCPv6 Two-Message Exchange

Two-message exchange applies to a network where only one DHCPv6 server is available. A DHCPv6 client multicasts a Solicit packet to locate the DHCPv6 server that can allocate addresses and configuration parameters. After receiving the Solicit packet, the DHCPv6 server responds with a Reply packet carrying addresses and configuration parameters allocated to the DHCPv6 client.

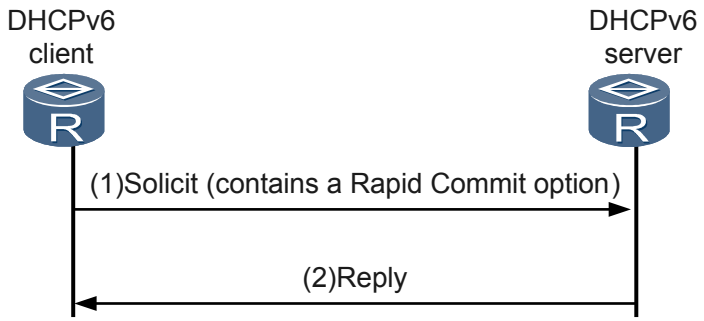
This packet exchange improves address allocation efficiency. On the network where multiple DHCPv6 servers are available, multiple DHCPv6 servers can allocate addresses to DHCPv6 clients and respond with Reply packets. The DHCPv6 clients, however, use the addresses and configuration parameters allocated by one DHCPv6 server. To prevent the preceding situation, the administrator can configure only one DHCPv6 server to support two-message exchange.

NOTE

- If a DHCPv6 server is configured with two-message exchange and the Solicit packet from a DHCPv6 client contains the Rapid Commit option, the DHCPv6 server allocates IPv6 addresses and configuration parameters in two-message exchange mode.
- If a DHCPv6 server does not support fast address allocation, the DHCPv6 server allocates IPv6 addresses and other network configuration parameters to clients using four-message exchange.

Figure 4-4 shows the process of address allocation using two-message exchange.

Figure 4-4 Process of address allocation using two-message exchange



The process of address allocation using two-message exchange is as follows:

Copyright © Huawei Technologies Co., Ltd.

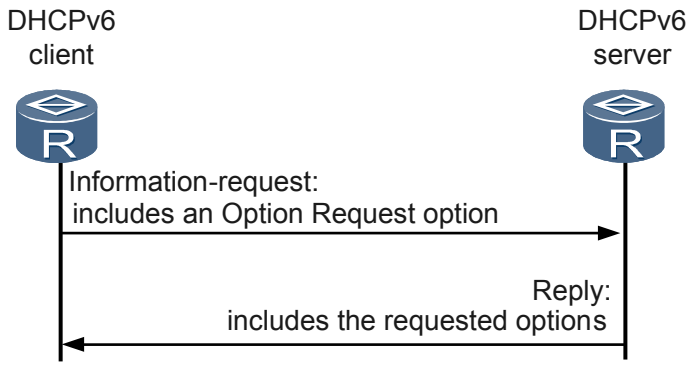
1. A DHCPv6 client sends a Solicit packet carrying the Rapid Commit option, indicating that the DHCPv6 client requires fast address allocation and network configuration parameters from a DHCPv6 server.
2. DHCPv6 server receives the Solicit message, it will processed as follows:
   - If the DHCPv6 server supports fast address allocation, it returns a Reply packet and allocates IPv6 addresses and other network configuration parameters to the DHCPv6 client.
   - If the DHCPv6 server does not support fast address allocation, the DHCPv6 server uses four-message exchange to allocate IPv6 addresses, prefixes, and other network configuration parameters.

DHCPv6 Stateless Autoconfiguration

The IPv6 node obtains network configuration parameters (including configuration parameters of DNS, SIP, and SNTP servers, without IPv6 addresses) through DHCPv6 stateless autoconfiguration.

Figure 4-5 shows the working process of DHCPv6 stateless autoconfiguration.

Figure 4-5 Working process of DHCPv6 stateless autoconfiguration



The working process of DHCPv6 stateless autoconfiguration is as follows:

1. A DHCPv6 client multicasts an Information-Request packet with the Option Request option to DHCPv6 servers. The Option Request option specifies the configuration parameters that the DHCPv6 client needs to obtain from a DHCPv6 server.
2. After receiving the Information-Request packet, the DHCPv6 server sends a Reply packet to the client in unicast mode. The Reply packet carries the allocated network configuration parameters.
3. The DHCPv6 client performs stateless autoconfiguration based on parameters carried in the Reply packet.

4.2.4 Working Principle of DHCPv6 PD

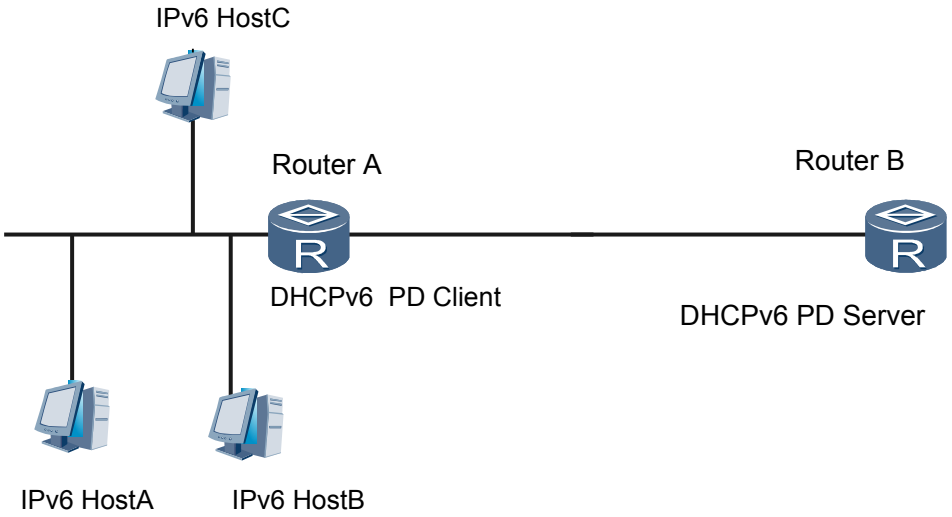
DHCPv6 prefix delegation (PD) is a prefix allocation mechanism proposed by Cisco and defined in RFC 3633. On a layered network, IPv6 addresses of different layers are configured manually. Manually configured IPv6 addresses have poor extensibility and cannot be planned and managed in a centralized manner.

Copyright © Huawei Technologies Co., Ltd.

The DHCPv6 PD mechanism allows a downstream router to request IPv6 prefixes from the upstream router and an upstream router to assign requested prefixes for the downstream router. In this way, you do not need to configure IPv6 prefixes for user-side links on the downstream router. The downstream router divides the obtained prefix (the length of the obtained prefix is smaller than 64 bits) into 64-bit prefix of subnet segments and sends a Route Advertisement (RA) packet on the link that IPv6 hosts directly connect to. This enables hosts to automatically configure addresses, completing IPv6 network deployment.

Figure 4-6 shows the working process of DHCPv6 PD.

Figure 4-6 Working principle of DHCPv6 PD



The process of DHCPv6 PD using four-message exchange is as follows:

1. A DHCPv6 PD client sends a Solicit packet, requesting an IPv6 address prefix from a DHCPv6 PD server.
2. If the DHCPv6 PD server does not support fast address allocation, the DHCPv6 PD server returns an Advertise packet containing the allocated address prefixes regardless of whether the Solicit packet contains the Rapid Commit option.
3. If receiving Advertise packets from multiple DHCPv6 PD servers, the DHCPv6 PD client selects the DHCPv6 PD server with the highest priority and sends a Request packet to this DHCPv6 PD server to request address prefixes.
4. The DHCPv6 PD server responds with a Reply packet to assign an IPv6 address prefix to the DHCPv6 PD client.

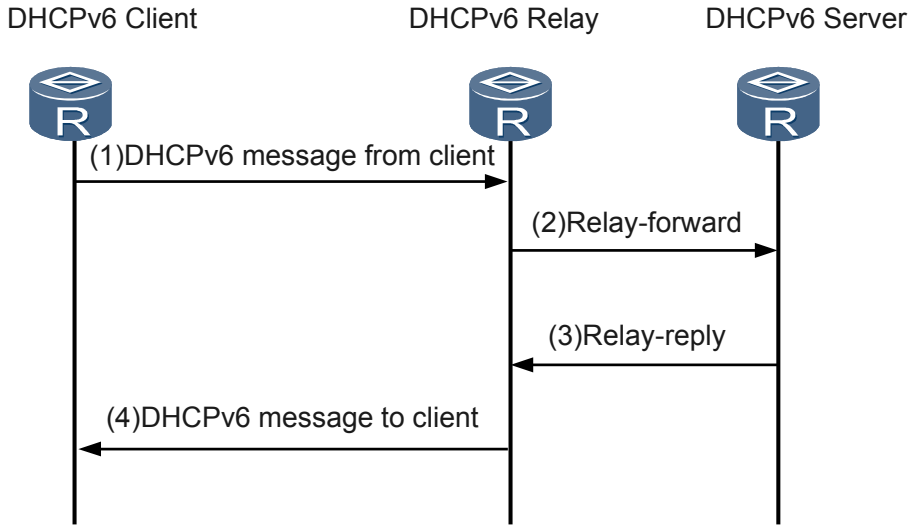
DHCPv6 PD also supports two-message exchange using packets carrying the Rapid Commit option. For details, see **DHCPv6 Two-Message Exchange**

4.2.5 Working Principle of the DHCPv6 Relay Agent

Figure 4-7 shows the working process of a DHCPv6 relay agent. A DHCPv6 client sends packets to a DHCPv6 server through a DHCPv6 relay agent to obtain IPv6 addresses, prefixes, and other network configuration parameters, such as IPv6 addresses of DNS servers.

Copyright © Huawei Technologies Co., Ltd.

Figure 4-7 Working principle of a DHCPv6 relay agent



The working process of a DHCPv6 relay agent is as follows:

1. A DHCPv6 client sends a Request packet to all the DHCPv6 servers and the DHCPv6 relay agent using multicast address FF02::1:2.
2. A DHCPv6 relay agent processes packets in the following two ways:
   - If a DHCPv6 relay agent and a DHCPv6 client are located on the same link, that is, the DHCPv6 relay agent is the first-hop relay agent of the DHCPv6 client, the DHCPv6 relay agent is the IPv6 gateway of the DHCPv6 client. After receiving a packet from the DHCPv6 client, the DHCPv6 relay agent encapsulates the packet in the Relay Message option of a Relay-Forward packet. Then the DHCPv6 relay agent sends the Relay-forward packet to a DHCPv6 server or the next hop relay agent.
   - If the DHCPv6 relay agent and DHCPv6 client are on different links, the DHCPv6 relay agent receives Relay-Forward packets sent from other relay agents. The DHCPv6 relay agent constructs a new Relay-Forward packet and sends the packet to the DHCPv6 server or the next hop relay agent.
3. The DHCPv6 server parses the request of the DHCPv6 client in the Relay-Forward packet and selects IPv6 addresses and other network configuration parameters to construct a reply packet. Then the DHCPv6 server encapsulates the reply packet in the Relay Message option in a Relay-Reply packet and sends the Relay-reply packet to the DHCPv6 relay agent.
4. The DHCPv6 relay agent parses the reply packet of the DHCPv6 server in the Relay-Reply packet and forwards the reply packet to the DHCPv6 client.The DHCPv6 client selects a DHCPv6 server according to priorities of DHCPv6 servers in Advertise packets sent by DHCPv6 servers in response to Solicit packets.

4.2.6 IPv6 Address/Prefix Allocation and Lease Updating

IPv6 Address Allocation Sequence

The DHCPv6 server allocates an IPv6 address or prefix to a DHCPv6 client in the following sequence:

Copyright © Huawei Technologies Co., Ltd.

1. Select an IPv6 address pool.

   An IPv6 address pool must be bound to an interface of the DHCPv6 server. The DHCPv6 server assigns addresses and prefixes to DHCPv6 clients from the IPv6 address pool.

2. Select an IPv6 address or prefix.

   After the address pool is configured, the DHCPv6 server assigns IPv6 addresses or prefixes to DHCPv6 clients in the following procedures:

   a. If IPv6 addresses or prefixes have been specified in the address pool, these addresses and prefixes matching the client DUIDs are preferentially assigned to clients.

   b. If the IA option in the packet sent from the client carries valid addresses or prefixes, these addresses or prefixes are preferentially assigned to clients from the address pool. If these addresses or prefixes are unavailable in the address pool, other idle addresses or prefixes are assigned to clients. If the IPv6 prefix length exceeds the assigned length, the IPv6 prefix of the assigned length is assigned.

   c. Idle addresses and prefixes are assigned to clients from the address pool. Reserved addresses (For example, anycast addresses defined in RFC 2526), conflicted addresses, and used addresses cannot be assigned to clients.

   d. If no IPv6 address or prefix can be assigned, address or prefix allocation fails.

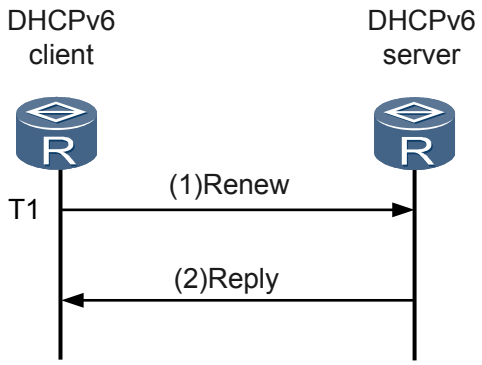
DHCPv6 Address Lease Updating

The addresses allocated by DHCPv6 servers to DHCPv6 clients have leases. A lease is composed of the lifetime (including the preferred lifetime and valid lifetime) and lease extension time (T1 and T2 in an IA). After the valid lifetime of an address is reached, a DHCPv6 client can no longer use this address. Before the valid lifetime is reached, a DHCPv6 client needs to update the address lease if it needs to continue to use this address.

To extend the valid lifetime and preferred lifetime for the addresses associated with an IA, a DHCPv6 client sends a Renew packet to the DHCPv6 server at T1. The IA option in the Renew packet carries the addresses whose leases need to be extended. If the DHCPv6 client does not receive a response packet, it sends a Rebind packet at T2 to the DHCPv6 server to continue to extend the address lease.

Figure 4-8 shows the process of updating the address lease at T1.

Figure 4-8 Process of updating the address lease at T1



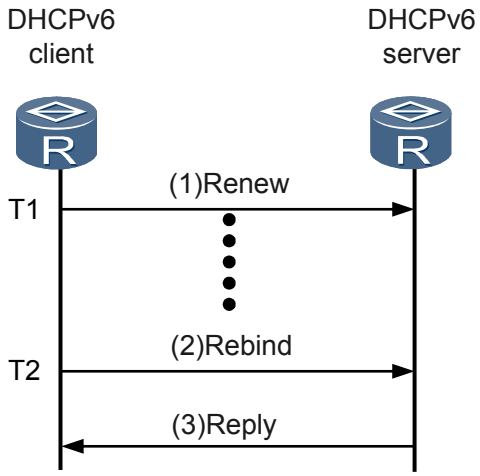
The process of updating the address lease at T1 is as follows:

Copyright © Huawei Technologies Co., Ltd.

1. A DHCPv6 client sends a Renew packet to request to update the address lease at T1 (the recommended value of T1 is half the preferred lifetime).
2. A DHCPv6 server responds with a Reply packet.
   - If the DHCPv6 client can continue to use the address, the DHCPv6 server responds with a Reply packet indicating that the address lease is extended successfully. In addition, the DHCPv6 server informs the DHCPv6 client that the address lease is updated successfully.
   - If the DHCPv6 client cannot use the address, the DHCPv6 server responds with a Reply packet indicating that address lease extension fails. In addition, the DHCPv6 server informs the DHCPv6 client that the DHCPv6 client cannot obtain a new address lease.

Figure 4-9 shows the process of updating the address lease at T2.

Figure 4-9 Process of updating the address lease at T2



The process of updating the address lease at T2 is as follows:

1. A DHCPv6 client sends a Renew packet to request to update the address lease at T1, but does not receive a response packet from a DHCPv6 server.
2. The DHCPv6 client multicasts a Rebind packet to all the DHCPv6 servers to request them to update the address lease at T2 (the recommended value of T2 is 0.8 times the preferred lifetime).
3. A DHCPv6 server responds with a Reply packet.
   - If the DHCPv6 client can continue to use the address, the DHCPv6 server responds with a Reply packet indicating that the address lease is extended successfully. In addition, the DHCPv6 server informs the DHCPv6 client that the address or prefix lease is updated successfully.
   - If the DHCPv6 client cannot use the address, the DHCPv6 server responds with a Reply packet indicating that address lease extension fails. In addition, the DHCPv6 server informs the DHCPv6 client that the DHCPv6 client cannot obtain a new address lease.

   NOTE

   If the DHCPv6 client does not receive a response packet from the DHCPv6 server, the DHCPv6 client stops using this address after the valid lifetime is reached.

Copyright © Huawei Technologies Co., Ltd.

IP Address Reservation

The DHCPv6 server supports reserved IPv6 addresses that cannot be dynamically allocated. For example, an IPv6 address can be reserved for a DNS server.

4.3 References

Table 1 RFCs related to DHCPv6 features

Document	Description	Remarks
RFC2460	Internet Protocol, Version 6 (IPv6) Specification	-
RFC3315	Dynamic Host Configuration Protocol for IPv6 (DHCPv6)	-
RFC3736	Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6	-
RFC2462	IPv6 Stateless Address Autoconfiguration	-

Copyright © Huawei Technologies Co., Ltd.

5 IPv6

About This Chapter

5.1 Introduction to IPv6

5.2 Principles

5.3 References

Copyright © Huawei Technologies Co., Ltd.

5.1 Introduction to IPv6

Definition

Internet Protocol version 6 (IPv6), also called IP Next Generation (IPng), is a second-generation network layer protocol. It was designed by the Internet Engineering Task Force (IETF) as an upgraded version of Internet Protocol version 4 (IPv4).

Purpose

IPv4 is the widely used Internet protocol. During initial development of the Internet, IPv4 rapidly developed because of its simplicity, ease of implementation, and good interoperability. However, as the Internet rapidly develops, deficiency in IPv4 design becomes obvious. To overcome the deficiency, IPv6 emerges. IPv6 has the following advantages over IPv4.

Table 5-1 Comparisons between IPv6 and IPv4

Item	Deficiency in IPv4	Advantage of IPv6
Address space	An IPv4 address is 32 bits long. A maximum of 4.3 billion IPv4 addresses can be provided. Actually, less than 4.3 billion addresses are available, and IPv4 address resources are allocated unevenly. USA address resources account for almost half of the global address space, with barely enough addresses left for Europe, and still fewer for the Asia-Pacific area. Furthermore, the development of mobile IP and broadband technologies still requires more IP addresses. Currently, IPv4 addresses are being exhausted. There are several solutions to IPv4 address exhaustion. Classless Inter-domain Routing (CIDR) and Network Address Translator (NAT) are two such solutions. CIDR and NAT, however, have their disadvantages and unsolvable problems, which helped encourage the development of IPv6.	An IPv6 address is 128 bits long. A 128-bit address structure allows for 2^{128} (4.3 billion x 4.3 billion x 4.3 billion x 4.3 billion) possible addresses. The biggest advantage of IPv6 is its almost infinite address space.

Copyright © Huawei Technologies Co., Ltd.

Item	Deficiency in IPv4	Advantage of IPv6
Packet format	The IPv4 packet header carries the Options field, including security, timestamp, and record route options. The variable length of the Options field makes the IPv4 packet header length range from 20 bytes to 60 bytes. IPv4 packets with the Options field often need to be forwarded by intermediate routers, so many router resources are occupied. Therefore, these IPv4 packets are seldom used in practice.	Compared with the IPv4 packet header, the IPv6 packet header does not carry IHL, identifier, flag, fragment offset, header checksum, option, and paddiing fields but carries the flow label field. This facilitates IPv6 packet processing and improves processing efficiency. To support various options without changing the existing packet format, the Extension Header information field is added to the IPv6 packet header. This improves IPv6 flexibility.
Autoconfiguration and readdressing	An IPv4 address is 32 bits long, and IPv4 addresses are allocated unevenly. IP addresses often need to be reallocated during network expansion or replanning. Address autoconfiguration and readdressing are required to simplify address maintenance. Currently, IPv4 depends on the Dynamic Host Configuration Protocol (DHCP) to provide address autoconfiguration and readdressing.	IPv6 provides address autoconfiguration to allow hosts to automatically discover networks and obtain IPv6 addresses. This improves network manageability.
Route summarization	Many non-contiguous IPv4 addresses are allocated, so routes cannot be summarized effectively due to incorrect IPv4 address allocation and planning. The increasingly large routing table consumes a lot of memory and affects forwarding efficiency. Device manufacturers have to keep upgrading routers to improve route addressing and forwarding performance.	A huge address space allows for the hierarchical network design in IPv6. The hierarchical network design facilitates route summarization and improves forwarding efficiency.
End-to-end security support	Security is not fully considered in the design of IPv4. Therefore, the original IPv4 framework does not support end-to-end security.	IPv6 supports IP Security (IPSec) authentication and encryption at the network layer, so it provides end-to-end security.
Quality of Service (QoS) support	The increasing popularity of network conferences, network telephones, and network TVs requires better QoS to ensure real-time forwarding of these voice, data, and video services. However, IPv4 has no native mechanism to support QoS.	IPv6 has the Flow Label field, which guarantees QoS for voice, data, and video services.

Copyright © Huawei Technologies Co., Ltd.

Item	Deficiency in IPv4	Advantage of IPv6
Mobility	As the Internet develops, mobile IPv4 experiences some problems, such as triangle routing and source address filtering.	IPv6 has the native capability to support mobility. Compared to mobile IPv4, mobile IPv6 uses the neighbor discovery function to discover a foreign network and obtain a care-of address without using any foreign agent. The mobile node and peer node can communicate using the routing header and destination options header. This function solves the problems of triangle routing and source address filtering in mobile IPv4. Mobile IPv6 improves mobile communication efficiency and is transparent to the application layer.

5.2 Principles

5.2.1 IPv6 Addresses

IPv6 Address Formats

An IPv6 address is 128 bits long. It is written as eight groups of four hexadecimal digits (0 to 9, A to F), where each group is separated by a colon (:). For example, 2031:0000:130F:0000:0000:09C0:876A:130B is a valid IPv6 address. This IPv6 address format is the preferred format.

For convenience, IPv6 provides the compressed format. The following uses IPv6 address 2031:0000:130F:0000:0000:09C0:876A:130B as an example to describe the compressed format:

- Any zeros at the beginning of a group can be omitted. Then the given example becomes 2031:0:130F:0:0:9C0:876A:130B.
- A double colon (::) can be used in an IPv6 address when two or more consecutive groups contain all zeros. Then the given example can be written as 2031:0:130F::9C0:876A:130B.

NOTE

An IPv6 address can contain only one double colon (::). Otherwise, a computer cannot determine the number of zeros in a group when restoring the compressed address to the original 128-bit address.

IPv6 Address Structure

An IPv6 address has two parts:

- Network prefix: corresponds to the network ID of an IPv4 address. It is of *n* bits.
- Interface identifier (interface ID): corresponds to the host ID of an IPv4 address. It is of 128-*n* bits.

Copyright © Huawei Technologies Co., Ltd.

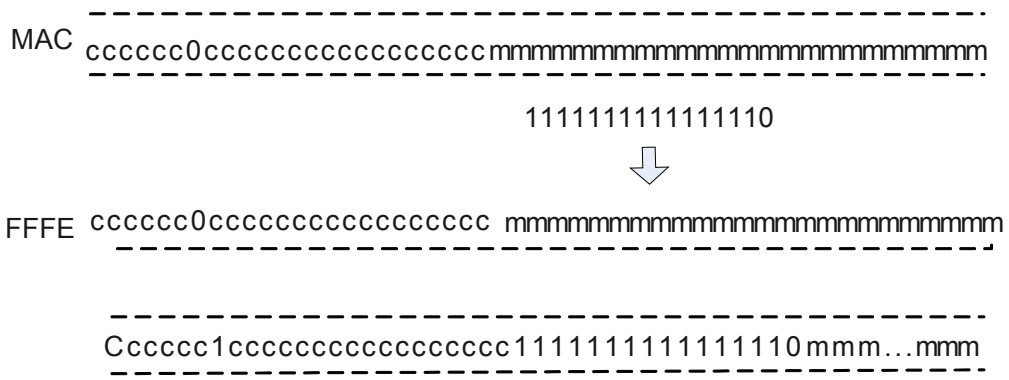
NOTE

If the first 3 bits of an IPv6 unicast address are not 000, the interface ID must be of 64 bits. If the first 3 bits are 000, there is no such limitation.

The interface ID can be manually configured, generated through the system software, or generated in IEEE 64-bit Extended Unique Identifier (EUI-64) format. It is most common to generate the interface ID in EUI-64 format.

IEEE EUI-64 standards convert an interface MAC address into an IPv6 interface ID. As shown in **Figure 5-1**, if a 48-bit MAC address is used as an interface ID, the first 24 bits (expressed by c) is a vendor identifier, and the last 24 bits (expressed by m) is an extension identifier. If the higher seventh bit is 0, the MAC address is locally unique. During conversion, EUI-64 inserts FFFE between the vendor identifier and extension identifier of the MAC address, and then the higher seventh bit 0 is changed to 1 to indicate that the interface ID is globally unique.

Figure 5-1 EUI-64 format



For example, if the MAC address is 000E-0C82-C4D4, the interface ID is 020E: 0CFF:FE82:C4D4 after the conversion.

The method for converting MAC addresses into IPv6 interface IDs reduces the configuration workload. When stateless address autoconfiguration is used, you only need an IPv6 network prefix before obtaining an IPv6 address. The defect of this method is that an IPv6 address can be easily calculated based on a MAC address.

IPv6 Address Types

IPv6 addresses are classified into unicast, anycast, and multicast addresses. Compared to IPv4, IPv6 has no broadcast address, uses multicast addresses as broadcast addresses, and introduces a new address type anycast address.

IPv6 Unicast Address

An IPv6 unicast address identifies an interface. Each interface belongs to a node. Therefore, the IPv6 unicast address of any interface on a node can identify the node. Packets sent to an IPv6 unicast address are delivered to the interface identified by the unicast address.

IPv6 defines multiple unicast addresses, including unspecified address, loopback address, global unicast address, link-local address, and unique local address.

- Unspecified address

Copyright © Huawei Technologies Co., Ltd.

An IPv6 unspecified address is 0:0:0:0:0:0:0:0/128 or ::/128, indicating that an interface or a node does not have an IP address. It can be used as the source IP address of some packets, such as Neighbor Solicitation (NS) message in duplicate address detection. Routers do not forward the packets with the source IP address as an unspecified address.

- Loopback address

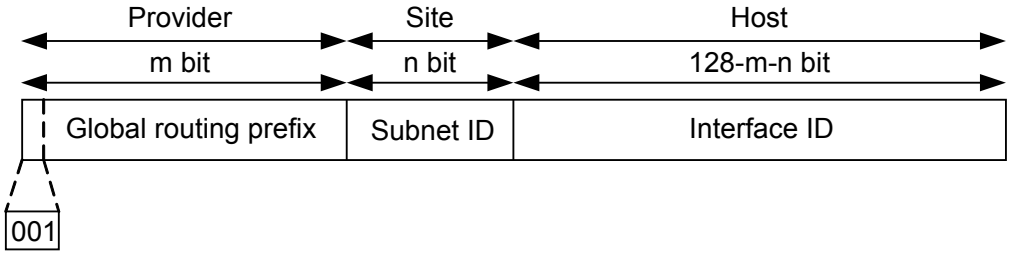
  An IPv6 loopback address is 0:0:0:0:0:0:0:1/128 or ::1/128. Similar to IPv4 loopback address 127.0.0.1, IPv6 loopback address is used when a node needs to send IPv6 packets to itself. This IPv6 loopback address is usually used as the IP address of a virtual interface (a loopback interface for example). The loopback address cannot be used as the source or destination IP address of packets that need to be forwarded.

- Global unicast address

  An IPv6 global unicast address is an IPv6 address with a global unicast prefix, which is similar to an IPv4 public address. IPv6 global unicast addresses support route prefix summarization, helping limit the number of global routing entries.

  A global unicast address consists of a global routing prefix, subnet ID, and interface ID, as shown in **Figure 5-2**.

  Figure 5-2 Global unicast address format



  Global routing prefix: is assigned by a service provider to an organization. A global routing prefix is of at least 48 bits. Currently, the first 3 bits of all the assigned global routing prefixes are 001.

  Subnet ID: is used by organizations to construct a local network (site). There are a maximum of 64 bits for both the global routing prefix and subnet ID. It is similar to an IPv4 subnet number.

  Interface ID: identifies a device (host).

- Link-local address

  Link-local addresses are used only in communication between nodes on the same local link. A link-local address uses a link-local prefix FE80::/10 as the first 10 bits (1111111010 in binary) and an interface ID as the last 64 bits.

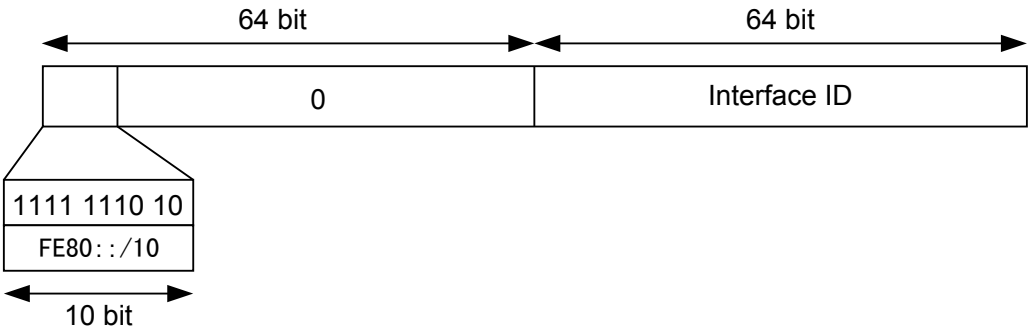
  When IPv6 runs on a node, each interface of the node is automatically assigned a link-local address that consists of a fixed prefix and an interface ID in EUI-64 format. This mechanism enables two IPv6 nodes on the same link to communicate without any configuration. Therefore, link-local addresses are widely used in neighbor discovery and stateless address configuration.

  Routers do not forward IPv6 packets with the link-local address as a source or destination address to devices on different links.

  Figure 5-3 shows the link-local address format.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-3 Link-local address format



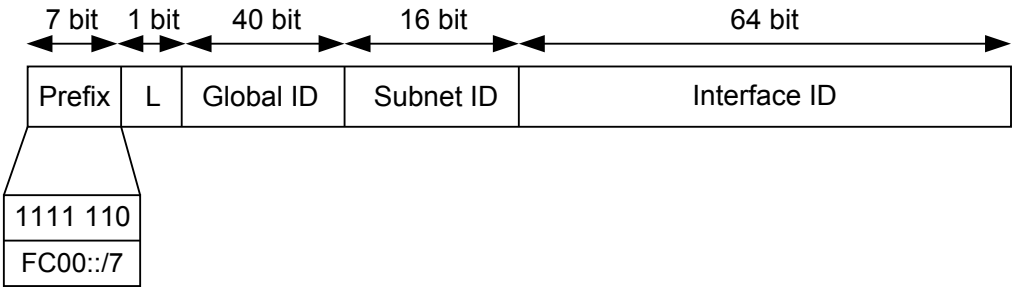
- Unique local address

  Unique local addresses are used only within a site. Site-local addresses are deprecated in RFC 3879 and replaced by unique local addresses in RFC 4193.

  Unique local addresses are similar to IPv4 private addresses. Any organization that does not obtain a global unicast address from a service provider can use a unique local address. Unique local addresses are routable only within a local network but not the Internet.

  Figure 5-4 shows the unique local address format.

Figure 5-4 Unique local address format



Prefix: is fixed as FC00::/7.

L: is set to 1 if the address is valid within a local network. The value 0 is reserved for future expansion.

Global ID: indicates a globally unique prefix, which is pseudo-randomly allocated (for details, see RFC 4193).

Subnet ID: identifies a subnet within the site.

Interface ID: identifies an interface.

A unique local address has the following characteristics:

- Has a globally unique prefix. The prefix is pseudo-randomly allocated and has a high probability of uniqueness.
- Allows private connections between sites without creating address conflicts.
- Has a well-known prefix (FC00::/7) that allows for easy route filtering at site boundaries.

Copyright © Huawei Technologies Co., Ltd.

- Does not conflict with any other addresses if it is leaked outside of the site through routing.
- Functions as a global unicast address to applications.
- Is independent of the Internet Service Provider (ISP).

IPv6 Multicast Address

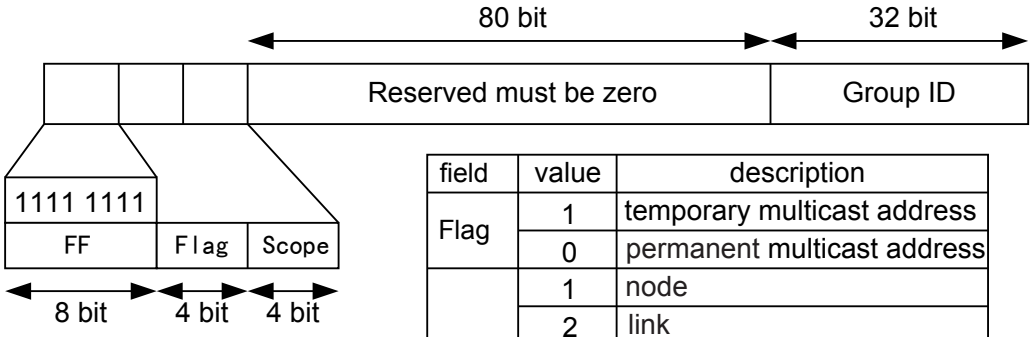
Like an IPv4 multicast address, an IPv6 multicast address identifies a group of interfaces, which usually belong to different nodes. A node may belong to any number of multicast groups. Packets sent to an IPv6 multicast address are delivered to all the interfaces identified by the multicast address.

An IPv6 multicast address is composed of a prefix, flag, scope, and group ID (global ID):

- Prefix: is fixed as FF00::/8 (1111 1111).
- Flag: is 4 bits long. The high-order 3 bits are reserved and must be set to 0s. The last bit 0 indicates a permanently-assigned (well-known) multicast address allocated by the Internet Assigned Numbers Authority (IANA). The last bit 1 indicates a non-permanently-assigned (transient) multicast address.
- Scope: is 4 bits long. It limits the scope where multicast data flows are sent on the network. **Figure 5-5** shows the field values and meanings.
- Group ID (global ID): is 112 bits long. It identifies a multicast group. RFC 2373 does not define all the 112 bits as a group ID but recommends using the low-order 32 bits as the group ID and setting all the remaining 80 bits to 0s. In this case, each multicast group ID maps to a unique Ethernet multicast MAC address (for details, see RFC 2464).

Figure 5-5 shows the IPv6 multicast address format.

Figure 5-5 IPv6 multicast address format



field	value	description
Flag	1	temporary multicast address
	0	permanent multicast address
Scope	1	node
	2	link
	4	management
	5	site
	8	organization
	E	global
	the rest	unsigned or reserved

- Solicited-node multicast address

  A solicited-node multicast address is generated using an IPv6 unicast or anycast address of a node. When a node has an IPv6 unicast or anycast address, a solicited-node multicast address is generated for the node, and the node joins the multicast group that corresponds to the IPv6 unicast or anycast address. A unicast or anycast address corresponds to a

Copyright © Huawei Technologies Co., Ltd.

solicited-node multicast address, which is often used in neighbor discovery and duplicate address detection.

IPv6 does not support broadcast addresses or Address Resolution Protocol (ARP). In IPv6, Neighbor Solicitation (NS) packets are used to resolve IP addresses to MAC addresses. When a node needs to resolve an IPv6 address to a MAC address, it sends an NS packet in which the destination IP address is the solicited-node multicast address corresponding to the IPv6 address.

The solicited-node multicast address consists of the prefix FF02::1:FF00:0/104 and the last 24 bits of the corresponding unicast address.

IPv6 Anycast Address

An anycast address identifies a group of network interfaces, which usually belong to different nodes. Packets sent to an anycast address are delivered to the nearest interface that is identified by the anycast address, depending on the routing protocols.

Anycast addresses are designed to implement the redundancy and load balancing functions when multiple hosts or nodes are provided with the same services. Currently, a unicast address is assigned to more than one interface to make a unicast address become an anycast address. When a unicast address is assigned to multiple hosts or nodes, the sender cannot determine which device can receive the sent data packets with the destination IP address as the anycast address, if there are multiple routes to the anycast address. This depends on the routing protocols running on the network. Anycast addresses are used in stateless applications, such as Domain Name Service (DNS).

IPv6 anycast addresses are allocated from the unicast address space. Anycast addresses are used in mobile IPv6 applications. Anycast prefix (2002:c058:6301::) is also used in IPv6-to-IPv4 relay.

NOTE

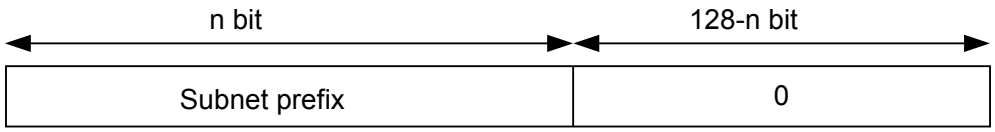
IPv6 anycast addresses can be assigned only to routers but not hosts. Anycast addresses cannot be used as the source IP addresses of IPv6 packets.

- Subnet-router anycast address

  A subnet-router anycast address is predefined in RFC 3513. Packets sent to a subnet-router anycast address are delivered to the nearest router on the subnet identified by the anycast address, depending on the routing protocols. All routers must support subnet-router anycast addresses. A subnet-router anycast address is used when a node needs to communicate with any of the routers on the subnet identified by the anycast address. For example, a mobile node needs to communicate with one of the mobile agents on the home subnet.

  In a subnet-router anycast address, the n-bit subnet prefix identifies a subnet and the remaining bits are padded with 0s. **Figure 5-6** shows the subnet-router anycast address format.

Figure 5-6 Subnet-router anycast address format



Copyright © Huawei Technologies Co., Ltd.

5.2.2 IPv6 Packet Format

An IPv6 packet has three parts: an IPv6 basic header, one or more IPv6 extension headers, and an upper-layer protocol data unit (PDU).

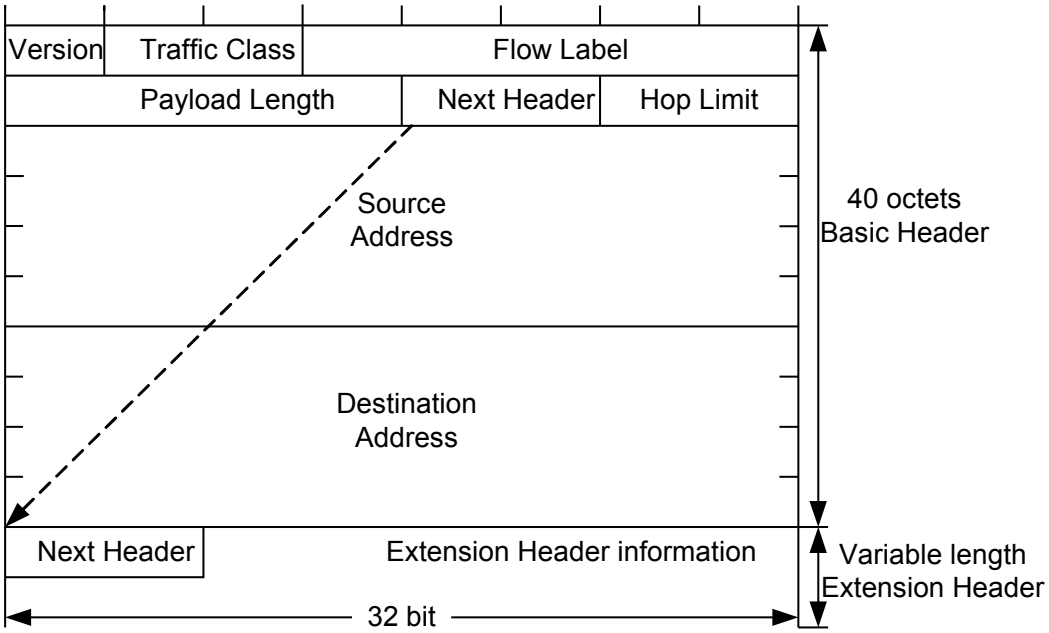
An upper-layer PDU is composed of the upper-layer protocol header and its payload such as an ICMPv6 packet, a TCP packet, or a UDP packet.

IPv6 Basic Header

An IPv6 basic header is fixed as 40 bytes long and has eight fields. Each IPv6 packet must have an IPv6 basic header. The IPv6 basic header provides basic packet forwarding information and will be parsed by all routers on the forwarding path.

Figure 5-7 shows the IPv6 basic header.

Figure 5-7 IPv6 basic header



An IPv6 basic header contains the following fields:

- Version: is 4 bits long. In IPv6, the Version field value is 6.
- Traffic Class: is 8 bits long. It indicates the class or priority of an IPv6 packet. The Traffic Class field is similar to the TOS field in an IPv4 packet and is mainly used in QoS control.
- Flow Label: is 20 bits long. This field is added in IPv6 to differentiate traffic. A flow label and source IP address identify a data flow. Intermediate network devices can effectively differentiate data flows based on this field.
- Payload Length: is 16 bits long, which indicates the length of the IPv6 payload. The payload is the rest of the IPv6 packet following this basic header, including the extension header and upper-layer PDU. This field indicates only the payload with the maximum length of 65535 bytes. If the payload length exceeds 65535 bytes, the field is set to 0. The payload length is expressed by the Jumbo Payload option in the Hop-by-Hop Options header.

Copyright © Huawei Technologies Co., Ltd.

- Next Header: is 8 bits long. This field identifies the type of the first extension header that follows the IPv6 basic header or the protocol type in the upper-layer PDU.
- Hop Limit: is 8 bits long. This field is similar to the Time to Live field in an IPv4 packet, defining the maximum number of hops that an IP packet can pass through. The field value is decremented by 1 by each device that forwards the IP packet. When the field value becomes 0, the packet is discarded.
- Source Address: is 128 bits long, which indicates the address of the packet originator.
- Destination Address: is 128 bits long, which indicates the address of the packet recipient.

Compared with the IPv4 packet header, the IPv6 packet header does not carry IHL, identifier, flag, fragment offset, header checksum, option, and paddiing fields but carries the flow label field. This facilitates IPv6 packet processing and improves processing efficiency. To support various options without changing the existing packet format, the Extension Header information field is added to the IPv6 packet header. This improves IPv6 flexibility. The following describes IPv6 extension headers.

IPv6 Extension Header

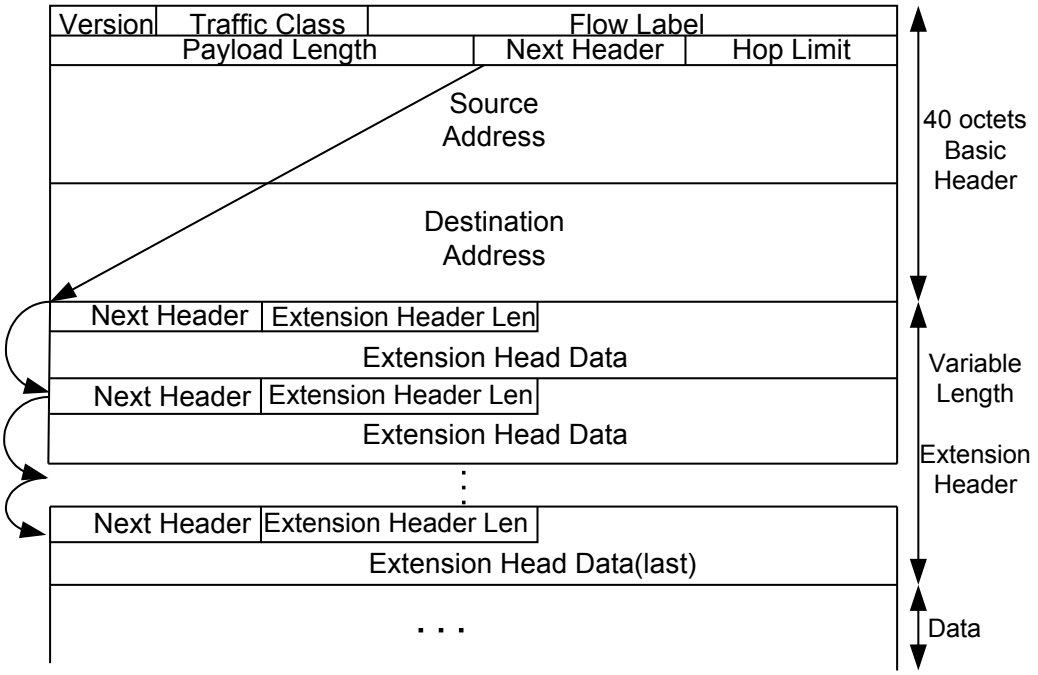
An IPv4 packet header has an optional field (Options), which includes security, timestamp, and record route options. The variable length of the Options field makes the IPv4 packet header length range from 20 bytes to 60 bytes. When routers forward IPv4 packets with the Options field, many resources need to be used. Therefore, these IPv4 packets are rarely used in practice.

To improve packet processing efficiency, IPv6 uses extension headers to replace the Options field in the IPv4 header. Extension headers are placed between the IPv6 basic header and upper-layer PDU. An IPv6 packet may carry zero, one, or more extension headers. The sender of a packet adds one or more extension headers to the packet only when the sender requests other routers or the destination router to perform special handling. Unlike IPv4, IPv6 has variable-length extension headers, which are not limited to 40 bytes. This facilitates further extension. To improve extension header processing efficiency and transport protocol performance, IPv6 requires that the extension header length be an integer multiple of 8 bytes.

When multiple extension headers are used, the Next Header field of an extension header indicates the type of the next header following this extension header. As shown in **Figure 5-8**, the Next Header field in the IPv6 basic header indicates the type of the first extension header, and the Next Header field in the first extension header indicates the type of the next extension header. If the next extension header does not exist, the Next Header field indicates the upper-layer protocol type. **Figure 5-8** shows the IPv6 extension header format.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-8 IPv6 extension header format



An IPv6 extension header contains the following fields:

- Next Header: is 8 bits long. It is similar to the Next Header field in the IPv6 basic header, indicating the type of the next extension header (if existing) or the upper-layer protocol type.
- Extension Header Len: is 8 bits long, which indicates the extension header length excluding the Next Header field.
- Extension Head Data: is of variable length. It includes a series of options and the padding field.

RFC 2460 defines six IPv6 extension headers: Hop-by-Hop Options header, Destination Options header, Routing header, Fragment header, Authentication header, and Encapsulating Security Payload header.

Copyright © Huawei Technologies Co., Ltd.

Table 5-2 IPv6 extension headers

Header Type	Next Header Field Value	Description
Hop-by-Hop Options header	0	This header carries information that must be examined by every node along the delivery path of a packet. This header is used in the following applications: ● Jumbo payload (the payload length exceeds 65535 bytes) ● Prompting routers to check this option before the routers forward packets. ● Resource Reservation Protocol (RSVP)
Destination Options header	60	This header carries information that needs to be examined only by the destination node of a packet. Currently, this header is used in mobile IPv6.
Routing header	43	Similar to the Loose Source and Record Route option in IPv4, this header is used by an IPv6 source node to specify the intermediate nodes that a packet must pass through on the way to the destination of the packet.
Fragment header	44	Like IPv4 packets, IPv6 packets to be forwarded cannot exceed the MTU. When the packet length exceeds the MTU, the packet needs to be fragmented. In IPv6, the Fragment header is used by an IPv6 source node to send a packet larger than the MTU.
Authentication header	51	This header is used in IPSec to provide data origin authentication, data integrity check, and packet anti-replay. It also protects some fields in the IPv6 basic header.
Encapsulating Security Payload header	50	Similar to the Authentication header, this header is used in IPSec to provide data origin authentication, data integrity check, packet anti-replay, and IPv6 packet encryption.

Conventions on IPv6 extension headers

When more than one extension header is used in the same packet, the headers must be listed in the following order:

- IPv6 basic header
- Hop-by-Hop Options header
- Destination Options header
- Routing header

Copyright © Huawei Technologies Co., Ltd.

- Fragment header
- Authentication header
- Encapsulating Security Payload header
- Destination Options header (for options to be processed only by the final destination of the packet)
- Upper-layer header

Intermediate routers determine whether to process extension headers according to the Next Header field value in the IPv6 basic header. Not all extension headers need to be examined and processed by intermediate routers.

Each extension header can only occur once in an IPv6 packet, except for the Destination Options header. The Destination Options header may occur at most twice (once before a Routing header and once before the upper-layer header).

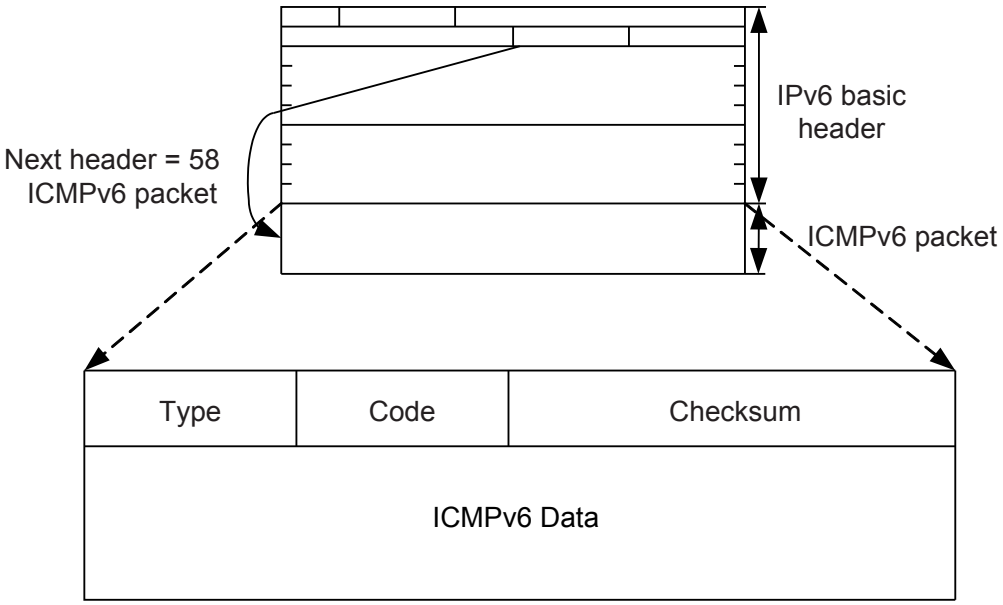
5.2.3 ICMPv6

The Internet Control Message Protocol version 6 (ICMPv6) is one of the basic IPv6 protocols.

In IPv4, ICMP reports IP packet forwarding information and errors to the source node. ICMP defines certain messages such as Destination Unreachable, Packet Too Big, Time Exceeded, and Echo Request or Echo Reply to facilitate fault diagnosis and information management. In addition to the common functions provided by ICMPv4, ICMPv6 provides mechanisms such as Neighbor Discovery (ID), stateless address configuration including duplicate address detection, and Path Maximum Transmission Unit (PMTU) discovery.

The protocol number of ICMPv6, namely, the value of the Next Header field in an IPv6 packet is 58. **Figure 5-9** shows the ICMPv6 packet format.

Figure 5-9 Format of an ICMPv6 packet



Each field is described as follows:

Copyright © Huawei Technologies Co., Ltd.

- Type: specifies the message type. Values 0 to 127 indicate the error message type, and values 128 to 255 indicate the informational message type.
- Code: indicates a specific message type.
- Checksum: indicates the checksum of an ICMPv6 packet.

Classification of ICMPv6 Error Messages

Error messages report errors generated during IPv6 packet forwarding. ICMPv6 error messages are classified into the following four types:

- Destination Unreachable message

  During IPv6 packet forwarding, if an IPv6 node detects that the destination address of a packet is unreachable, it sends an ICMPv6 Destination Unreachable message to the source node. Information about the causes for the error message is carried in the message.

  In an ICMPv6 Destination Unreachable message, the value of the Type field is 1. Based on different causes, the value of the Code field can be:

  - Code=0: No route to the destination device.
  - Code=1: Communication with the destination device is administratively prohibited.
  - Code=2: Not assigned.
  - Code=3: Destination IP address is unreachable.
  - Code=4: Destination port is unreachable.
- Packet Too Big message

  During IPv6 packet forwarding, if an IPv6 node detects that the size of a packet exceeds the link MTU of the outbound interface, it sends an ICMPv6 Packet Too Big message to the source node. The link MTU of the outbound interface is carried in the message. PMTU discovery is implemented based on Packet Too Big messages.

  In a Packet Too Big message, the value of the Type field is 2 and the value of the Code field is 0.
- Time Exceeded message

  During the transmission of IPv6 packets, when a router receives a packet with the hop limit being 0 or a router reduces the hop limit to 0, it sends an ICMPv6 Time Exceeded message to the source node. During the processing of a packet to be fragmented and reassembled, an ICMPv6 Time Exceeded message is also generated when the reassembly time is longer than the specified period.

  In a Time Exceeded message, the value of the Type field is 3. Based on different causes, the value of the Code field can be:

  - Code=0: Hop limit exceeded in packet transmission.
  - Code=1: Fragment reassembly timeout.
- Parameter Problem message

  When a destination node receives an IPv6 packet, it checks the validity of the packet. If an error is detected, it sends an ICMPv6 Parameter Problem message to the source node.

  In a Parameter Problem message, the value of the Type field is 4. Based on different causes, the value of the Code field can be:

  - Code=0: A field in the IPv6 basic header or extension header is incorrect.
  - Code=1: The Next Header field in the IPv6 basic header or extension header cannot be identified.

Copyright © Huawei Technologies Co., Ltd.

- Code=2: Unknown options exist in the extension header.

Classification of ICMPv6 Information Messages

ICMPv6 information messages provide the diagnosis and additional host functions such as Multicast Listener Discovery (MLD) and ND. Common ICMPv6 information messages include Ping messages that consist of Echo Request and Echo Reply messages.

- Echo Request messages: Echo Request messages are sent to destination nodes. After receiving an Echo Request message, the destination node responds with an Echo Reply message. In an Echo Request message, the value of the Type field is 128 and the value of the Code field is 0.
- Echo Reply messages: After receiving an Echo Request message, the destination node responds with an Echo Reply message. In an Echo Reply message, the value of the Type field is 129 and the value of the Code field is 0.

5.2.4 Neighbor Discovery

The Neighbor Discovery Protocol (NDP) is one important IPv6 basic protocol. It is an enhancement of the Address Resolution Protocol (ARP) and Internet Control Management Protocol (ICMP) router discovery. In addition to the function of ICMPv6 address resolution, NDP also provides the following functions: neighbor tracking, duplicate address detection, router discovery, and redirection.

Address Resolution

In IPv4, a host needs to obtain the link-layer address of the destination host through the ARP protocol for communication. Similar to IPv4, the IPv6 NDP protocol parses the IP address to obtain the link-layer address.

ARP packets are encapsulated in Ethernet packets. The Ethernet type value is 0x0806. ARP is defined as a protocol that runs between Layer 2 and Layer 3. ND is implemented through ICMPv6 packets. The IPv6 type value is 0x86dd. The Next Header value in the IPv6 header is 58, indicating that the packets are ICMPv6 packets. NDP packets are encapsulated in ICMPv6 packets. Therefore, NDP is taken as a Layer 3 protocol. Layer 3 address resolution brings the following advantages:

- Layer 3 address resolution enables Layer 2 devices to use the same address resolution protocol.
- Layer 3 security mechanisms such as IPSec are used to prevent address resolution attacks.
- Request packets are sent in multicast mode, reducing performance requirements on Layer 2 networks.

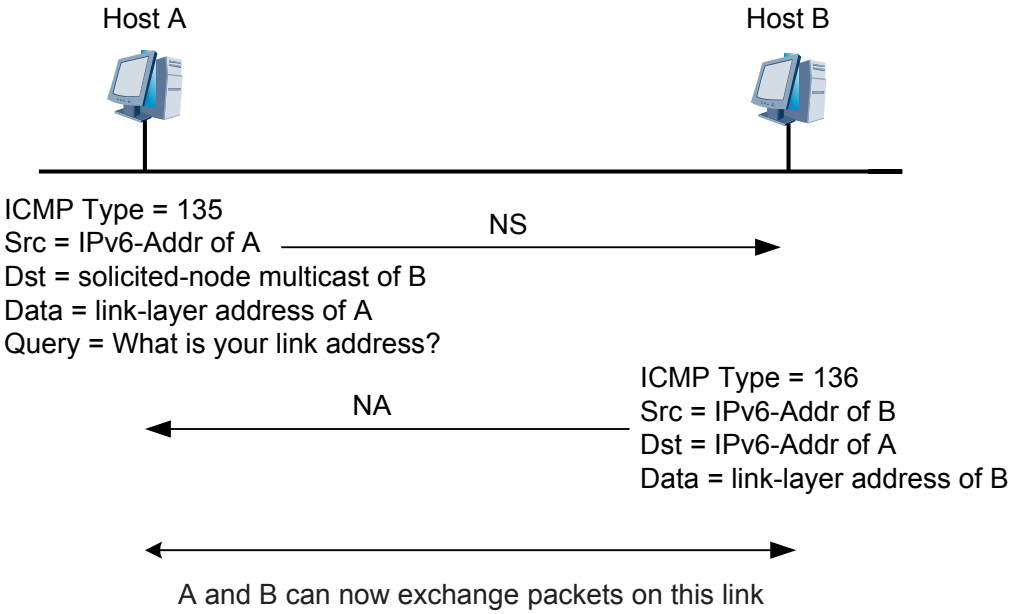
Neighbor Solicitation (NS) packets and Neighbor Advertisement (NA) packets are used during address resolution.

- In an NS packet, the value of the Type field is 135 and the value of the Code field is 0. An NS packet is similar to the ARP Request packet in IPv4.
- In an NA packet, the value of the Type field is 136 and the value of the Code field is 0. An NA packet is similar to the ARP Reply packet in IPv4.

Figure 5-10 shows the process of address resolution.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-10 IPv6 address resolution



Host A needs to parse the link-layer address of Host B before sending packets to Host B. Therefore, Host A sends an NS message on the network. In the NS message, the source IP address is the IPv6 address of Host A, and the destination IP address is the solicited-node multicast address of Host B. The destination IP address to be parsed is the IPv6 address of Host B. This indicates that Host A wants to know the link-layer address of Host B. The Options field in the NS message carries the link-layer address of Host A.

After receiving the NS message, Host B replies with an NA Reply message. In the NA reply message, the source address is the IPv6 address of Host B, and the destination address is the IPv6 address of Host A (the NS message is sent to Host A in unicast mode using the link-layer address of Host A). The Options field carries the link-layer address of Host B. This is the whole address resolution process.

Neighbor Tracking

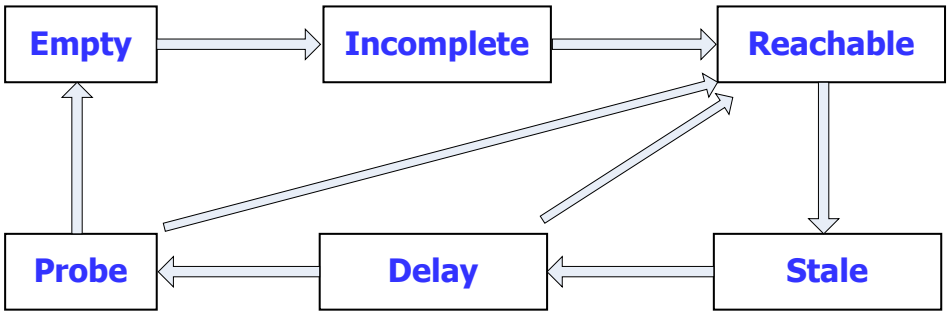
Communication with neighboring devices will be interrupted because of various reasons such as hardware fault and hot swapping of interface cards. If the destination address of a neighboring device becomes invalid, communication cannot be restored. If the path fails, communication can be restored. Therefore, nodes need to maintain the neighbor table to monitor the status of each neighboring device. A neighbor state can transit from one to another.

Five neighbor states are defined in RFC2461: Incomplete, Reachable, Stale, Delay, and Probe.

Figure 5-11 shows the transition of neighbor states.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-11 Neighbor state transition



The following example describes the neighbor state changes of node A during the first communication with node B.

1. Node A sends an NS message and generates a cache entry. The neighbor state of node A is Incomplete.
2. If node B replies with an NA message, the neighbor state of node A changes from Incomplete to Reachable; otherwise, the neighbor state changes from Incomplete to Empty after a certain period of time. Node A deletes this entry.
3. After the neighbor reachable time times out, the neighbor state changes from Reachable to Stale, indicating that whether the neighbor is reachable is unknown.
4. If node A in the Reachable state receives a non-NA Request message from node B, and the link-layer address of node B carried in the message is different from that learned by node A, the neighbor state of node A immediately goes to Stale.
5. If node A in the Stale state sends data to node B, the state of node A changes from Stale to Delay. Node A sends an NS Request message.
6. After a certain period of time, the neighbor state changes from Delay to Probe. During this time, if node A receives an NA Reply message, the neighbor state of node A changes to Reachable.
7. Node A in the Probe state sends unicast NS messages at the configured interval for several times. If node A receives a Reply message, the neighbor state of node A changes from Probe to Reachable; otherwise, the state changes to Empty. Node A deletes this entry.

Duplicate Address Detection

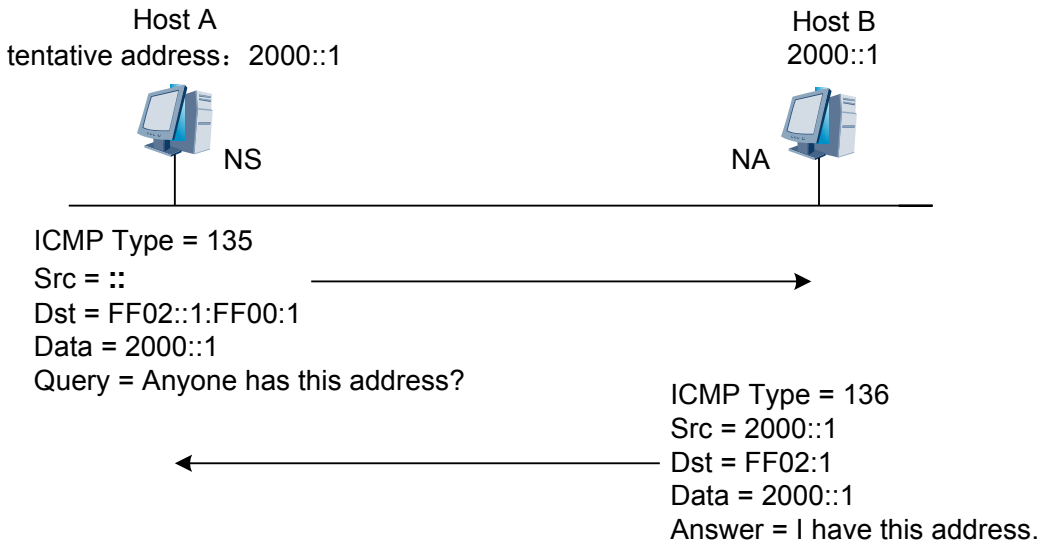
Before an IPv6 unicast address is assigned to an interface, duplicate address detection (DAD) is performed to check whether the address is used by another node. DAD is required if IP addresses are configured automatically. An IPv6 unicast address that is assigned to an interface but has not been verified by DAD is called a tentative address. An interface cannot use the tentative address for unicast communication but will join two multicast groups: ALL-nodes multicast group and Solicited-node multicast group.

IPv6 DAD is similar to IPv4 free ARP. A node sends an NS message that requests the tentative address as the destination address to the Solicited-node multicast group. If the node receives an NA Reply message, the tentative address is being used by another node. This node will not use this tentative address for communication.

Figure 5-12 shows the DAD working principle.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-12 DAD example



An IPv6 address 2000::1 is assigned to Host A as a tentative IPv6 address. To check the validity of 2000::1, Host A sends an NS message to the Solicited-node multicast group to which 2000::1 belongs. The NS message contains the requested address 2000::1. Since 2000::1 is not specified, the source address of the NS message is an unspecified address. After receiving the NS message, Host B processes the message in the following ways:

- If 2000::1 is one tentative address of Host B, Host B will not use this address as an interface address and not send the NA message.
- If 2000::1 is being used on Host B, Host B sends an NA message to the Solicited-node multicast group to which 2000::1 belongs. The NA message carries IP address 2000::1. Host A receives the message, finding that the tentative address is being used. Then, Host A abandons the address.

Router Discovery

Router discovery is used to locate a neighboring router and learn the address prefix and configuration parameters for address autoconfiguration.

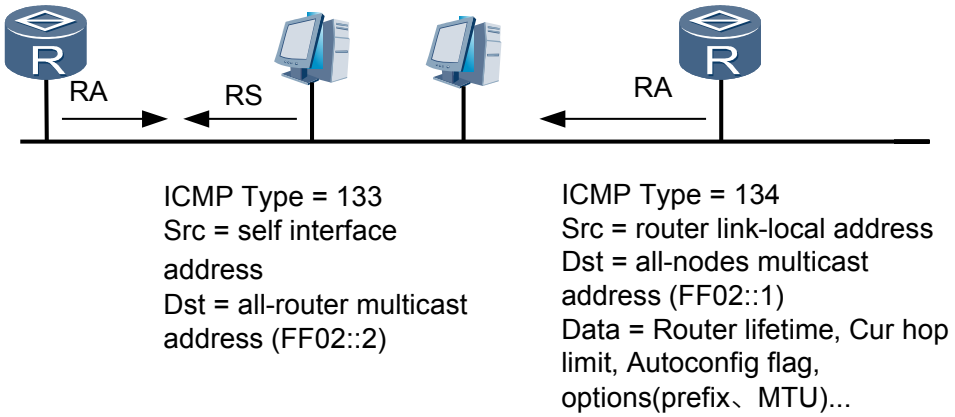
IPv6 supports stateless address autoconfiguration. Hosts obtain IPv6 prefixes and automatically generate interface IDs. Router Discovery is the basics for IPv6 address autoconfiguration and is implemented through the following two packets:

- Router Advertisement (RA) message: Each router periodically sends multicast RA messages that carry network prefixes and identifiers on the network to declare its existence to Layer 2 hosts and routers. An RA message has a value of 134 in the Type field.
- Router Solicitation (RS) message: After being connected to the network, a host immediately sends an RS message to obtain network prefixes. Routers on the network reply with an RA message. An RS message has a value of 133 in the Type field.

Figure 5-13 shows the router discovery function.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-13 Router discovery example



Address Autoconfiguration

IPv4 uses DHCP to automatically configure IP addresses and default gateways. This simplifies network management. The length of an IPv6 address is increased to 128 bits. Multiple terminal nodes require the function of automatic configuration. IPv6 allows both stateful and stateless address autoconfiguration. Stateless autoconfiguration enables hosts to automatically generate link-local addresses. Based on the prefixes in the RA message, hosts automatically configure global unicast addresses and obtain other information.

The process of IPv6 stateless autoconfiguration is as follows:

1. A host automatically configures the link-local address based on the interface ID.
2. The host sends an NS message for duplicate address detection.
3. If address conflict occurs, the host stops address autoconfiguration. Then addresses need to be configured manually.
4. If addresses do not conflict, the link-local address takes effect. The host is connected to the network and can communicate with the local node.
5. The host sends an RS message or receives RA messages routers periodically send.
6. The host obtains the IPv6 address based on the prefixes carried in the RA message and the configured interface ID specified by EUI-64.

Default Router Priority and Route Information Discovery

If multiple routers exist on the Internet where hosts reside, hosts need to select forwarding routers based on the destination address of the packet. In such a case, routers advertise default router priorities and route information, which allows hosts to select the optimal forwarding router based on the packet destination address.

The fields of default router priority and route information are defined in an RA message. These two fields enable hosts to select the optimal forwarding router.

After receiving an RA message that contains route information, hosts update their routing tables. When sending packets to other devices, hosts check the route in the routing table and select the optimal route.

When receiving an RA message that carries default router priorities, hosts update their default router lists. When sending packets to other devices, hosts check the router list to select the router with the highest priority to forward packets. If the selected router does not work, hosts select the router in descending order of priorities.

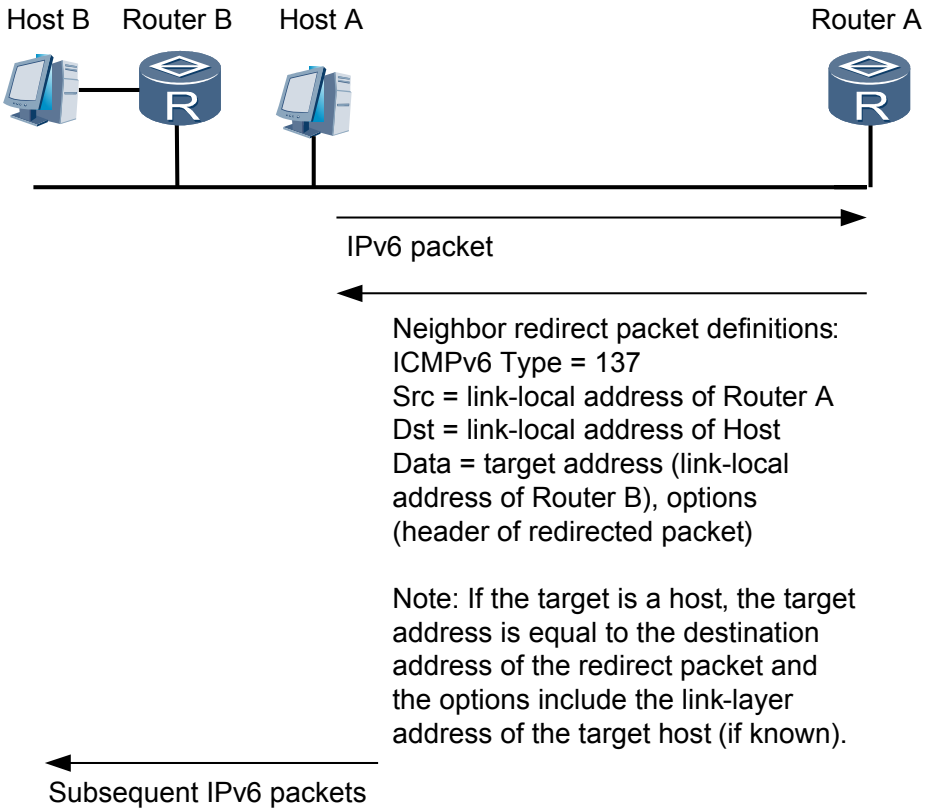
Copyright © Huawei Technologies Co., Ltd.

Redirection

To choose an optimal gateway router, the gateway router sends a Redirection message to notify the sender that packets can be sent from another gateway router. A Redirection message is contained in an ICMPv6 message. A Redirection message has the value of 137 in the Type field and carries a better next hop address and destination address of packets that need to be redirected.

Figure 5-14 shows the process of redirecting packets.

Figure 5-14 Packet redirection example



Host A needs to communicate with Host B. By default, packets sent from Host A to Host B are sent through Router A. After receiving packets from Host A, Router A finds that sending packets to Router B is much better. Router A sends a Redirection message to Host A to notify Host A that Router B is a better next hop address. The destination address of Host B is carried in the Redirection message. After receiving the Redirection message, Host A adds a host route to the default routing table. Packets sent to Host B will be directly sent to Router B.

A router sends a Redirection message in the following situations:

- The destination address of the packet is not a multicast address.
- Packets are not forwarded to the router through the route.
- After route calculation, the outbound interface of the next hop is the interface that receives the packets.
- The router finds that a better next hop IP address of the packet is on the same network segment as the source IP address of the packet.

Copyright © Huawei Technologies Co., Ltd.

- After checking the source address of the packet, the router finds a neighboring device in the neighbor entries that uses this address as the global unicast address or the link-local unicast address.

5.2.5 Path MTU

In IPv4, a packet needs to be fragmented if it is oversized. When the transit device receives from a source node a packet whose size exceeds the maximum transmission unit (MTU) of its outbound interface, the transit device fragments the packet before forwarding it to the destination node. In IPv6, however, packets are fragmented on the source node to reduce the pressure on the transit device. When an interface on the transit device receives a packet whose size exceeds the MTU, the transit device discards the packet and sends an ICMPv6 Packet Too Big message to the source node. The ICMPv6 Packet Too Big message contains the MTU value of the outbound interface. The source node fragments the packet based on the MTU and sends the packet again. This increases traffic overhead. The Path MTU Discovery (PMTUD) protocol dynamically discovers the MTU value of each link on the transmission path, reducing excessive traffic overhead.

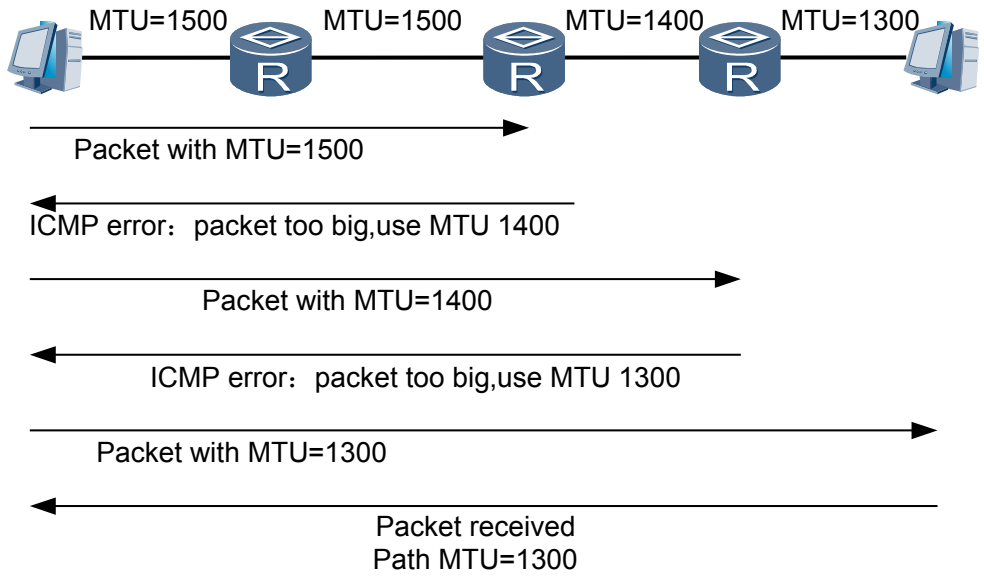
The PMTU protocol is implemented through ICMPv6 Packet Too Big messages. A source node first uses the MTU of its outbound interface as the PMTU and sends a probe packet. If a smaller PMTU exists on the transmission path, the transit device sends a Packet Too Big message to the source node. The Packet Too Big message contains the MTU value of the outbound interface on the transit device. After receiving the message, the source node changes the PMTU value to the received MTU value and sends packets based on the new MTU. This process is repeated until packets are sent to the destination address. Then the source node obtains the PMTU of the destination address.

NOTE

The switch supports the MTU setting on a VLANIF interface. Then packets sent by the protocol stack are fragmented based on the configured MTU. However, the hardware chip does not support the MTU setting, and the default MTU is 12K.

Figure 5-15 shows the process of PMTU discovery.

Figure 5-15 PMTU discovery



Copyright © Huawei Technologies Co., Ltd.

Packets are transmitted through four links. The MTU values of the four links are 1500, 1500, 1400, and 1300 bytes respectively. Before sending a packet, the source node fragments the packet based on PMTU 1500. When the packet is sent to the outbound interface with MTU 1400, the router returns a Packet Too Big message that carries MTU 1400. After receiving the message, the source node fragments the packet based on MTU 1400 and sends the fragmented packet again. When the packet is sent to the outbound interface with MTU 1300, the router returns another Packet Too Big message that carries MTU 1300. The source node receives the message and fragments the packet based on MTU 1300. In this way, the source node sends the packet to the destination address and discovers the PMTU of the transmission path.

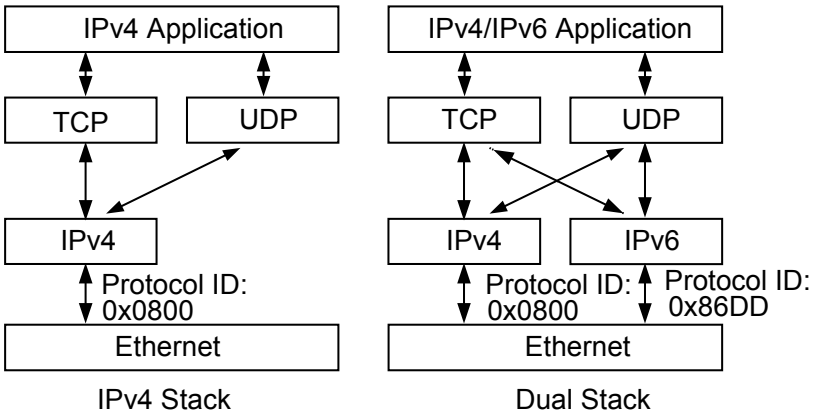
NOTE

IPv6 allows a minimum MTU of 1280 bytes. Therefore, the PMTU must be greater than 1280 bytes. PMTU of 1500 bytes is recommended.

5.2.6 Dual Protocol Stack

Dual protocol stack is a technology used for the transition from the IPv4 to IPv6 network. Nodes on a dual stack network support both IPv4 and IPv6 protocol stacks. A source node and a destination node use the same protocol stack. Network devices use protocol stacks to process and forward packets based on the protocol type of packets. You can implement a dual protocol stack on a unique device or a dual stack backbone network. On the dual stack backbone network, all devices must support both IPv4 and IPv6 protocol stacks. Interfaces connecting to the dual stack network must be configured with both IPv4 and IPv6 addresses. **Figure 5-16** shows the structures of a single protocol stack and a dual protocol stack.

Figure 5-16 Dual protocol stack



A dual protocol stack has the following advantages:

- Supported by multiple link protocols.

  Multiple link protocols, such as Ethernet, support dual protocol stacks. In **Figure 5-16**, the link protocol is Ethernet. In an Ethernet frame, if the value of the Protocol ID field is 0x0800, the network layer receives IPv4 packets. If the value of the Protocol ID field is 0x86DD, the network layer receives IPv6 packets.
- Supported by multiple applications.

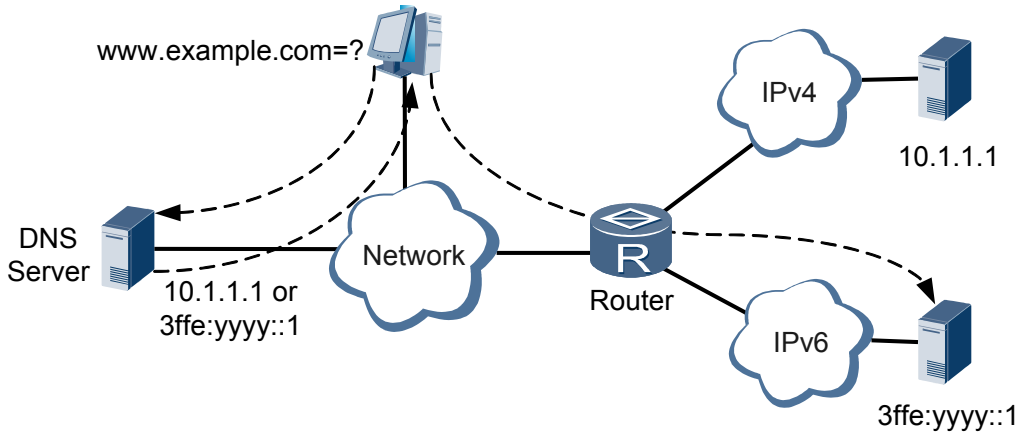
  Multiple applications, such as the DNS, FTP, and Telnet, support dual protocol stacks. The upper layer applications, such as the DNS, can use TCP or UDP as the transport layer

Copyright © Huawei Technologies Co., Ltd.

protocol. However, they prefer the IPv6 protocol stack rather than the IPv4 protocol stack as the network layer protocol.

Figure 5-17 shows a typical application of the dual IPv4/IPv6 protocol stack.

Figure 5-17 Networking diagram for applying a dual protocol stack



As shown in Figure 5-17, an application that supports dual protocol stack requests an IP address corresponding to the domain name **www.example.com** from the DNS server. As shown in the figure, a host sends a DNS request packet to the DNS server, requesting the IP address corresponding to the domain name **www.example.com**. The DNS server responds with the requested IP address. The IP address can be 10.1.1.1 or 3ffe:yyyy::1. If the host sends a class A query packet, it requests the IPv4 address from the DNS server. If the host sends a class AAAA query packet, it requests the IPv6 address from the DNS server.

Router in the figure supports the dual protocol stack. Router uses the IPv4 protocol stack to connect the host to the network server with the IPv4 address 10.1.1.1. Router uses the IPv6 protocol stack to connect the host to the network server with the IPv6 address 3ffe:yyyy::1.

5.2.7 IPv6 over IPv4 Tunnel

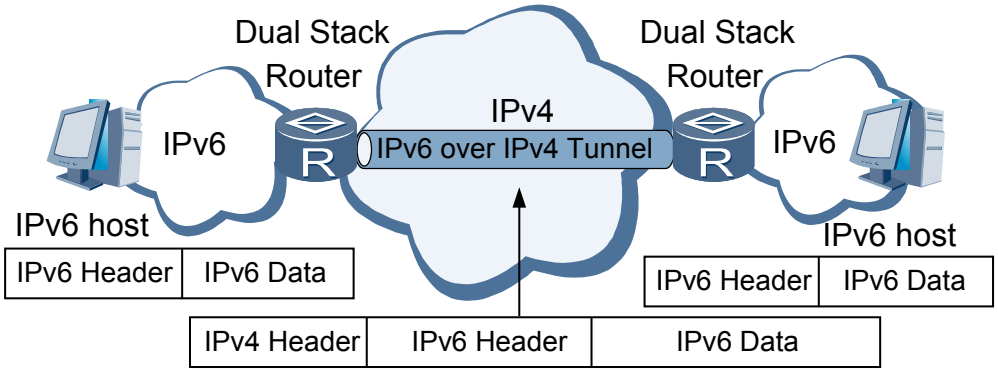
Tunnel is an encapsulation technology. Tunnel technology encapsulates packets of a network layer protocol as packets of another one for transmission. A tunnel is a virtual point-to-point (P2P) connection. It provides a path through which encapsulated packets are transmitted. Datagrams are encapsulated at one end and then decapsulated at the other end of the tunnel. Tunnel technology refers to the process that datagrams are encapsulated, transmitted, and decapsulated. It is of great importance for the transition from IPv4 to IPv6.

Exhaustion of IPv4 addresses brings an urgent demand for transition to IPv6. As IPv6 is not compatible with IPv4, you need to replace devices on the original IPv4 network. Replacing a large number of devices on the IPv4 network costs a lot and causes service interruption of the current network. Therefore, transition from IPv4 networks to IPv6 networks must be performed step by step. During the early transition, a large number of IPv4 networks have been deployed, whereas IPv6 networks are isolated sites over the world. You can create tunnels on the IPv4 networks to connect to IPv6 isolated sites. These tunnels are called IPv6 over IPv4 tunnels.

Figure 5-18 shows how to apply the IPv6 over IPv4 tunnel.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-18 Networking diagram for applying the IPv6 over IPv4 tunnel



1. On the border router, the dual IPv4/IPv6 protocol stack is enabled, and an IPv6 over IPv4 tunnel is configured.
2. After the border router receives a packet from the IPv6 network, the router appends an IPv4 header to the IPv6 packet to encapsulate the IPv6 packet as an IPv4 packet if the destination address of the IPv6 packet is not the router and the outbound interface of the next hop is the tunnel interface.
3. On the IPv4 network, the encapsulated packet is transmitted to the remote border router.
4. The remote border router decapsulates the packet, removes the IPv4 header, and sends the decapsulated IPv6 packet to the IPv6 network.

A tunnel is established when its start and end points are determined. You must manually configure an IPv4 address at the start point of an IPv6 over IPv4 tunnel. The IPv4 address at the end point of the tunnel can be determined manually or automatically. Based on the mode in which the end point IPv4 address is obtained, IPv6 over IPv4 tunnels are classified into manual tunnels and automatic tunnels.

- Manual tunnel: If a tunnel is created manually, a border router cannot automatically obtain an IPv4 address at the end point. You must manually configure an end point IPv4 address before packets can be transmitted to the remote border router.
- Automatic tunnel: If a tunnel is created automatically, a border router can automatically obtain an IPv4 address at the end point. The addresses of two interfaces on both ends of the tunnel are IPv6 addresses with IPv4 addresses embedded. The border router extracts IPv4 addresses from destination IPv6 addresses.

Manual Tunnel

Based on encapsulation modes of IPv6 packets, manual tunnels are classified into IPv6 over IPv4 manual tunnels and IPv6 over IPv4 Generic Routing Encapsulation (GRE) tunnels.

IPv6 over IPv4 Manual Tunnel

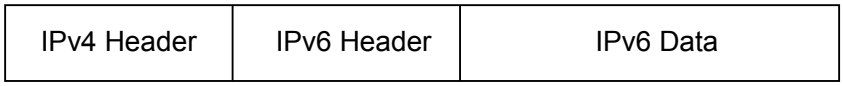
The border router uses the received IPv6 packet as the payload and encapsulates the IPv6 packet as an IPv4 packet. You must manually specify the source and destination addresses of a manual tunnel. A manual tunnel is a P2P connection. It can be created between two border routers to connect IPv4 isolated IPv6 sites, or created between a border router and a host to enable the host to access an IPv6 network. Hosts and border routers on both ends of a manual tunnel must support the IPv4/IPv6 dual protocol stack. Other devices only need to support a single protocol stack. If you create multiple IPv6 over IPv4 manual tunnels between one border router and multiple hosts,

Copyright © Huawei Technologies Co., Ltd.

the configuration workload is heavy. Therefore, an IPv6 over IPv4 manual tunnel is commonly created between two border routers to connect IPv6 networks.

Figure 5-19 shows the encapsulation format of an IPv6 over IPv4 packet.

Figure 5-19 Encapsulation format of an IPv6 over IPv4 packet



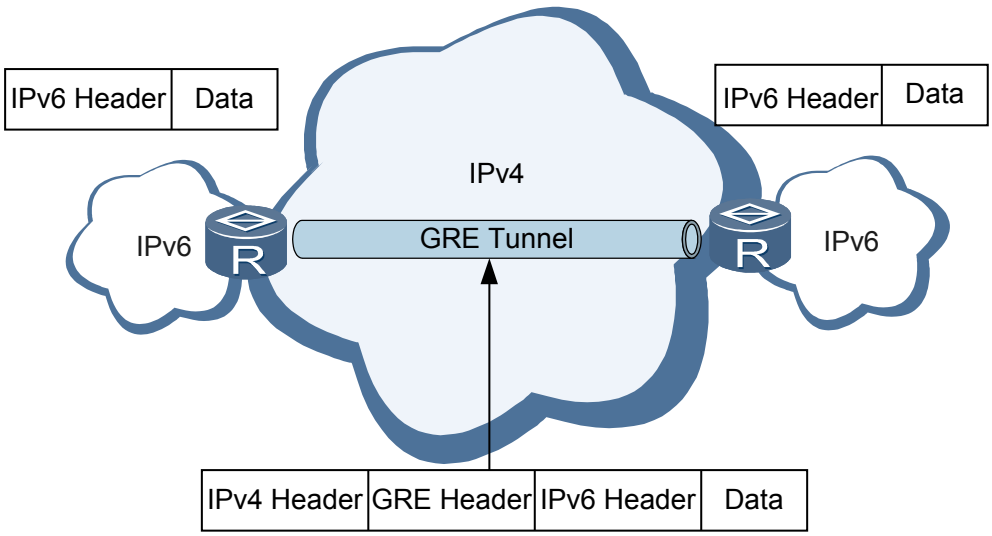
The forwarding mechanism of an IPv6 over IPv4 manual tunnel is as follows: After a border router receives an packet from the IPv6 network, it searches the destination address of the IPv6 packet in the routing and forwarding table. If the packet is forwarded from this virtual tunnel interface, the router encapsulates the packet based on the source and destination IPv4 addresses configured on the interface. The IPv6 packet is encapsulated as an IPv4 packet and processed by the IPv4 protocol stack. The encapsulated packet is forwarded through the IPv4 network to the remote end of the tunnel. After the border router on the remote end of the tunnel receives the encapsulated packet, it decapsulates the packet and processes the packet using the IPv6 protocol stack.

IPv6 over IPv4 GRE Tunnel

An IPv6 over IPv4 GRE tunnel uses the standard GRE tunnel technology to provide P2P connections. You must manually specify addresses for both ends of the tunnel. Any types of protocol packets that GRE supports can be encapsulated and transmitted through a GRE tunnel. The protocols may include IPv4, IPv6, Open Systems Interconnection (OSI), and Multiprotocol Label Switching (MPLS).

Figure 5-20 shows the encapsulation and transmission process on an IPv6 over IPv4 GRE tunnel.

Figure 5-20 IPv6 over IPv4 GRE tunnel



The forwarding mechanism of an IPv6 over IPv4 GRE tunnel is the same as that of an IPv6 over IPv4 manual tunnel. For details, see the *Feature Description - VPN*.

Copyright © Huawei Technologies Co., Ltd.

Automatic Tunnel

You only need to configure the start point of an automatic tunnel, and the device automatically obtains the end point of the tunnel. The tunnel interface uses a special form of IPv6 address with an IPv4 address embedded. The device obtains the IPv4 address from the destination IPv6 address and uses the IPv4 address as the end point address of the tunnel.

Based on the encapsulation modes of IPv6 packets, automatic tunnels are classified into IPv4-compatible IPv6 automatic tunnels, IPv6-to-IPv4 tunnels, and Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) tunnels.

IPv4-compatible IPv6 Automatic Tunnel

For an IPv4-compatible IPv6 automatic tunnel, the destination address contained in an IPv6 packet is an IPv4-compatible IPv6 address. The first 96 bits of an IPv4-compatible IPv6 address are all 0s and the last 32 bits are the IPv4 address. **Figure 5-21** shows the format of an IPv4-compatible IPv6 address.

Figure 5-21 IPv4-compatible IPv6 address

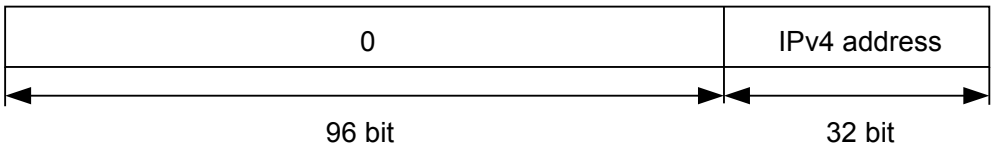
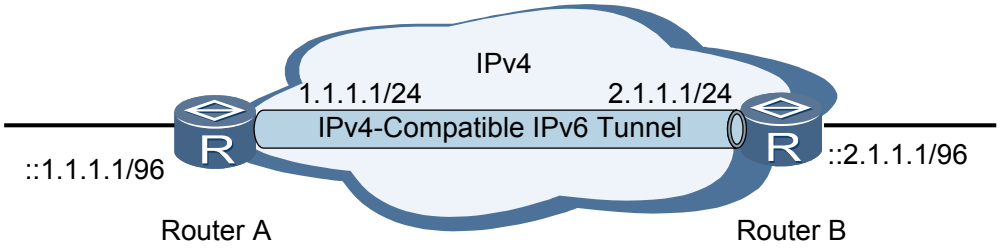


Figure 5-22 shows the forwarding mechanism of an IPv4-compatible IPv6 automatic tunnel.

Figure 5-22 Forwarding mechanism of an IPv4-compatible IPv6 automatic tunnel



After receiving an IPv6 packet, Router A searches the routing table for the destination address :: 2.1.1.1 and finds that the next hop address is a virtual tunnel interface address. Router A then encapsulates the IPv6 packet as an IPv4 address because the tunnel configured on Router A is an IPv4-compatible IPv6 automatic tunnel. The source address of the encapsulated IPv4 address is the start point address of the tunnel 1.1.1.1, and the destination address is 2.1.1.1, which is the last 32 bits of the IPv4-compatible IPv6 address. Router A sends the packet through the tunnel interface and forwards it on an IPv4 network to the destination address 2.1.1.1 (Router B). Router B receives the packet, obtains the IPv6 packet, and processes the IPv6 packet using the IPv6 protocol stack. Router B returns packets to Router A in the same way.

NOTE

If the IPv4 address contained in an IPv4-compatible IPv6 address is a broadcast address, multicast address, network broadcast address, subnet broadcast address of an outbound interface, address of all 0s, or loopback address, the IPv6 packet will be discarded.

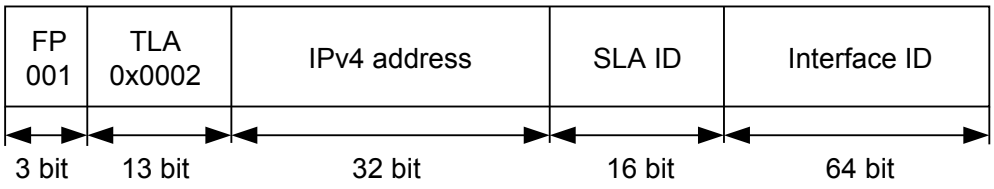
Copyright © Huawei Technologies Co., Ltd.

To deploy an IPv4-compatible IPv6 tunnel, each host must have a valid IP address, and hosts that communicate with each other must support dual protocol stacks and IPv4-compatible IPv6 tunnels. Therefore, it is unsuitable for large-scale networks. Currently, the IPv4-compatible IPv6 tunnel has been replaced by the IPv6-to-IPv4 tunnel.

IPv6-to-IPv4 Tunnel

An IPv6-to-IPv4 tunnel also uses an IPv4 address that is embedded in an IPv6 address. Unlike IPv4-compatible IPv6 tunnels, you can create IPv6-to-IPv4 tunnels between two routers, a router and a host, and two hosts. An IPv6-to-IPv4 address uses the IPv4 address as the network ID. **Figure 5-23** shows the format of an IPv6-to-IPv4 address.

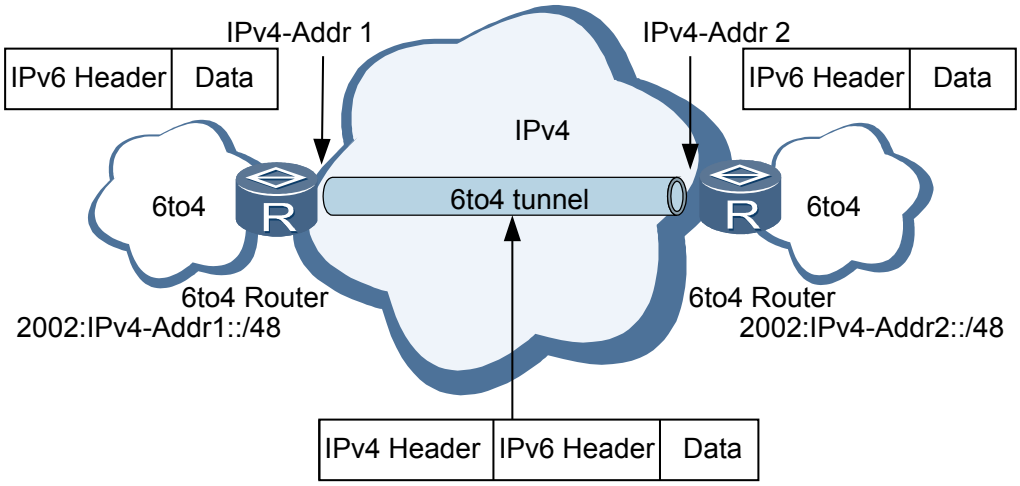
Figure 5-23 Format of an IPv6-to-IPv4 address



- FP: format prefix of a global unicast address. The value is 001.
- TLA ID: top level aggregation identifier. The value is 0x0002.
- SLA ID: site level aggregation identifier.

An IPv6-to-IPv4 address is expressed in the format of 2002::/16. An IPv6-to-IPv4 network is expressed as 2002:IPv4 address::/48. An IPv6-to-IPv4 address has a 64-bit prefix composed of 48-bit 2002:IPv4 address and 16-bit SLA. 2002:IPv4 address in the format of 2002:a.b.c.d is determined by the IPv4 address allocated to the router and the SLA is defined by the user. **Figure 5-24** shows the encapsulation and forwarding process of the IPv6-to-IPv4 tunnel. It is the same as that of the IPv4-compatible IPv6 automatic tunnel, and therefore it is not mentioned here.

Figure 5-24 Example of an IPv6-to-IPv4 tunnel (1)

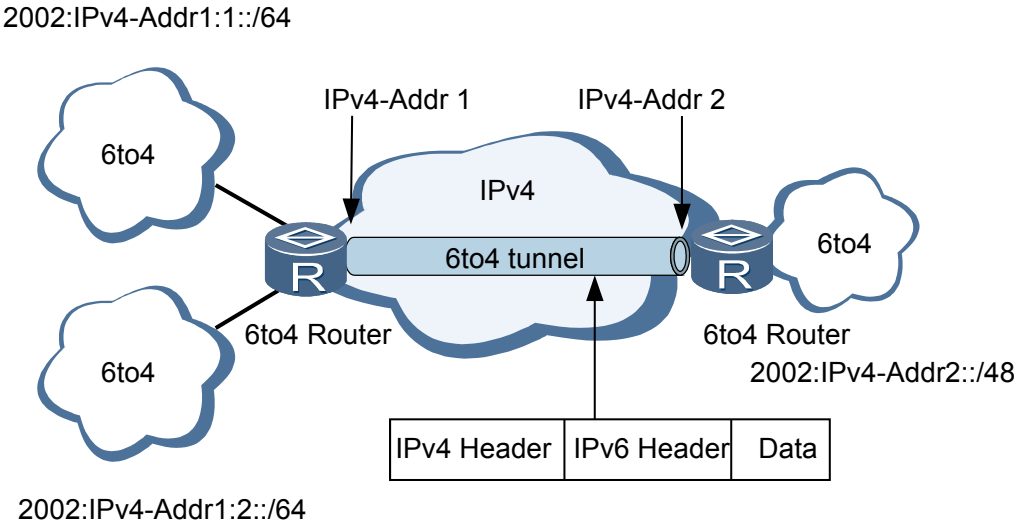


One IPv4 address can be used as the source address of only one IPv6-to-IPv4 tunnel. When a border router is connected to multiple IPv6-to-IPv4 networks that use the same IPv4 address as

Copyright © Huawei Technologies Co., Ltd.

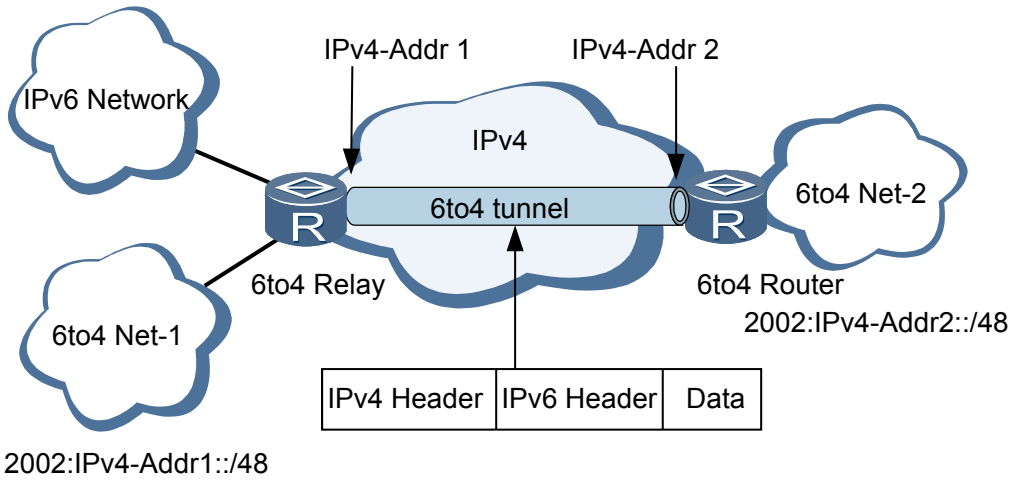
the source address of the tunnel, the IPv6-to-IPv4 networks share a tunnel and are identified by SLA ID in the IPv6-to-IPv4 address. **Figure 5-25** shows the case.

Figure 5-25 Example of an IPv6-to-IPv4 tunnel (2)



Backed by the advance of IPv6 networks, IPv6 hosts need to communicate with IPv4 hosts through IPv6-to-IPv4 networks. It can be implemented by deploying IPv6-to-IPv4 relays. When the destination address of an IPv6 packet forwarded through an IPv6-to-IPv4 tunnel is not an IPv6-to-IPv4 address, but the next hop address is an IPv6-to-IPv4 address, the next hop router is an IPv6-to-IPv4 relay. The device obtains the destination IPv4 address from the next hop IPv6-to-IPv4 address. **Figure 5-26** shows an IPv6-to-IPv4 relay.

Figure 5-26 IPv6-to-IPv4 relay



When hosts on IPv6-to-IPv4 network 2 want to communicate with hosts on the IPv6 network, configure the next hop address as the IPv6-to-IPv4 address of the IPv6-to-IPv4 relay on the border router. The IPv6-to-IPv4 address matches the source address of the IPv6-to-IPv4 tunnel. Packets sent from IPv6-to-IPv4 network 2 to the IPv6 network are sent to the IPv6-to-IPv4 relay

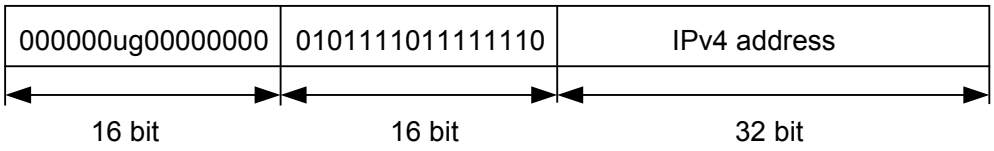
Copyright © Huawei Technologies Co., Ltd.

router according to the routing table. The IPv6-to-IPv4 relay router then forwards packets to the pure IPv6 network. When hosts on the IPv6 network send packets to IPv6-to-IPv4 network 2, the IPv6-to-IPv4 relay router appends IPv4 headers to the packets and forwards the packets to the destination addresses (IPv6-to-IPv4 addresses).

ISATAP Tunnel

ISATAP is another automatic tunnel technology. The ISATAP tunnel uses a special format of IPv6 address with an IPv4 address embedded. Different from the IPv6-to-IPv4 address that uses the IPv4 address as the network prefix, the ISATAP address uses the IPv4 address as the interface ID. **Figure 5-27** shows the format of the interface ID of an ISATAP address.

Figure 5-27 Format of the interface ID of an ISATAP address



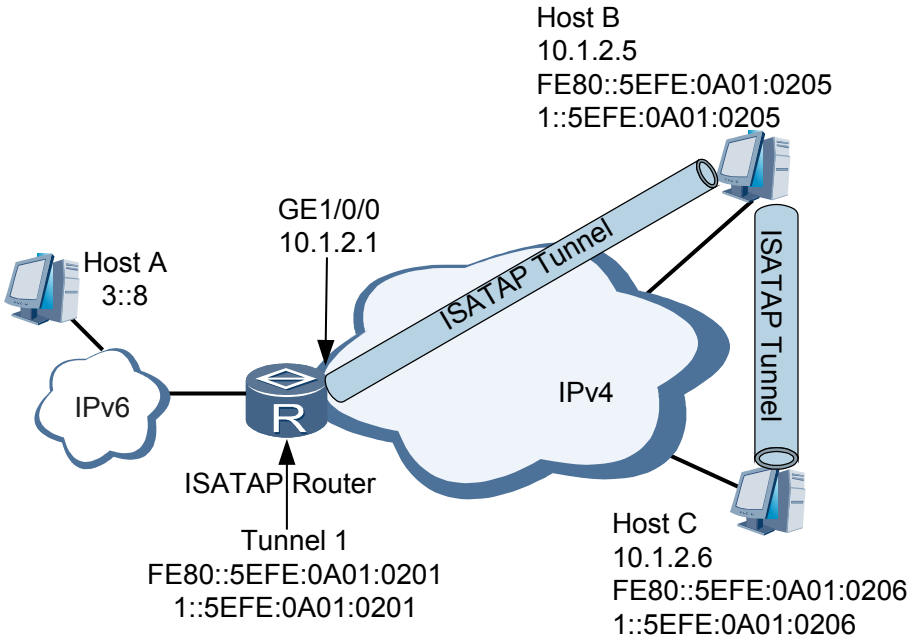
The "u" bit in the IPv4 address that is globally unique is set to 1. Otherwise, the "u" bit is set to 0. "g" is the individual/group bit. An ISATAP address contains an interface ID and it can be a global unicast address, link-local address, ULA address, or multicast address. The device obtains the first 64 bits of an ISATAP address by sending Request packets to the ISATAP router. Devices on both ends of the ISATAP tunnel run the Neighbor Discovery (ND) protocol. The ISATAP tunnel considers the IPv4 network as a non-broadcast multiple access (NBMA) network.

ISATAP allows IPv6 networks to be deployed within existing IPv4 networks. The deployment is simple and networks can be easily expanded. Therefore, ISATAP is suitable for transition of local sites. ISATAP supports local routing within IPv6 sites, global IPv6 routing domains, and automatic IPv6 tunnels. ISATAP can be used together with NAT to allow the use of an IPv4 address that is not globally unique within the site. Typically, an ISATAP tunnel is used within the site, and does not require a globally unique IPv4 address embedded.

Figure 5-28 shows a typical application of the ISATAP tunnel.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-28 Typical application of the ISATAP tunnel



As shown in **Figure 5-28**, Host B and Host C are located on an IPv4 network. They both support dual protocol stacks and have private IPv4 addresses. Perform the following operations to enable the ISATAP function on Host B and Host C:

1. Configure an ISATAP tunnel interface to generate an interface ID based on the IPv4 address.
2. Encapsulate a link-local IPv6 address based on the interface ID. When a host obtains the link-local IPv6 address, it can access the IPv6 network on the local link.
3. The host automatically obtains a global unicast IPv6 address and ULA address.
4. The host obtains an IPv4 address from the next hop IPv6 address as the destination address, and forwards packets through the tunnel interface to communicate with another IPv6 host. When the destination host is located on the same site as the source host, the next hop address is the address of the source host. When the destination host is not located on the local site, the next hop address is the address of the ISATAP router.

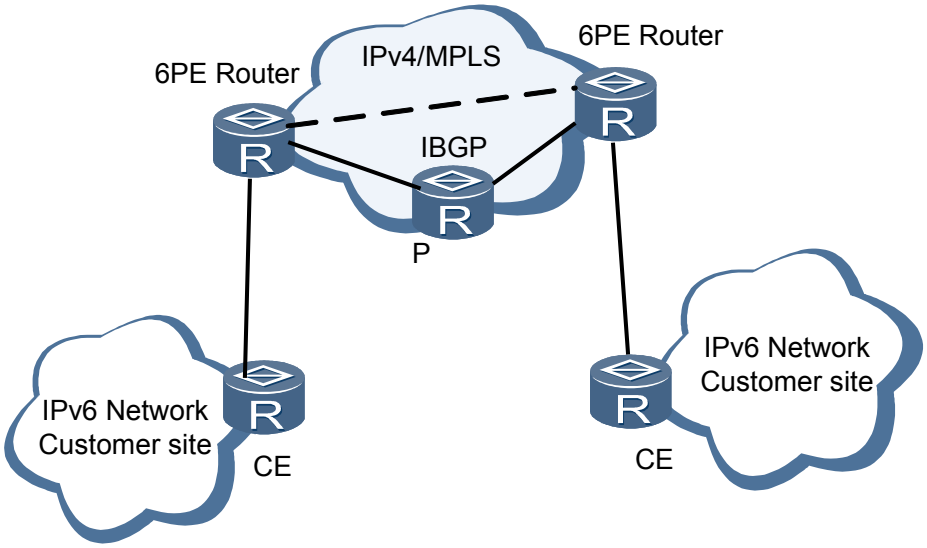
6PE

IPv6 Provider Edge (6PE) is a transition technology from the IPv4 to IPv6 network. With 6PE routers, Independent Service Providers (ISPs) can provide access services for the IPv6 networks of isolated users over the existing IPv4 backbone network. The 6PE router labels IPv6 routing information and floods the information onto the ISP's IPv4 backbone network through Internal Border Gateway Protocol (IBGP) sessions. The IPv6 packets are labeled before flowing into tunnels on the backbone network. The tunnels can be GRE tunnels or MPLS LSPs. To allow IPv6 packet exchange on IPv4/MPLS networks through MPLS, LSPs can just update the PE routers. Therefore, using the 6PE technology as an IPv6 transition mechanism is a cost-effective solution for ISPs.

Figure 5-29 shows the typical 6PE networking diagram.

Copyright © Huawei Technologies Co., Ltd.

Figure 5-29 Typical 6PE networking diagram



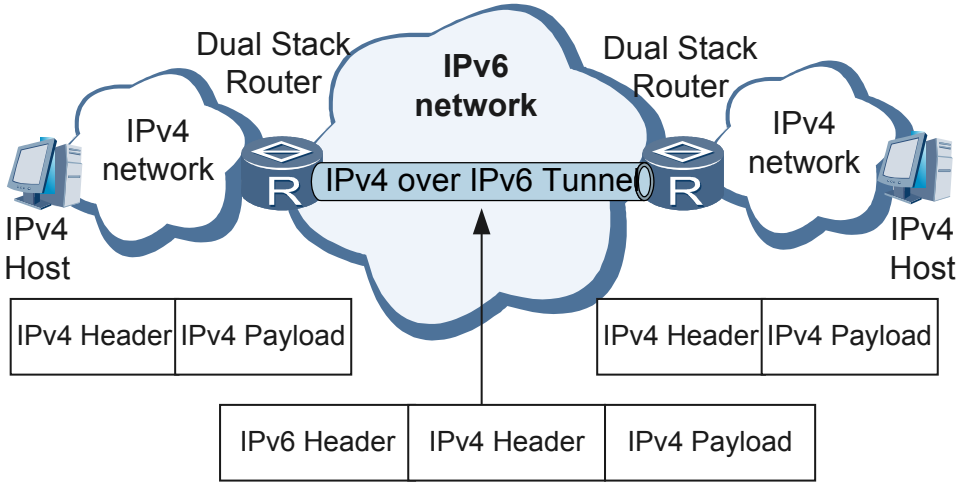
For details about 6PE, see the *Feature Description - MPLS*. The details are not described here.

5.2.8 IPv4 over IPv6 Tunnel

During the later transition from IPv4 networks to IPv6 networks, a large number of IPv6 networks are deployed. IPv4 networks, however, are isolated sites over the world. You can create tunnels on the IPv6 networks to connect IPv4 isolated sites so that IPv4 isolated sites can access other IPv4 networks through the IPv6 public network.

Figure 5-30 shows how to apply the IPv4 over IPv6 tunnel.

Figure 5-30 Networking diagram for applying the IPv4 over IPv6 tunnel



1. On the border router, the IPv4/IPv6 dual protocol stack is enabled and the IPv4 over IPv6 tunnel is configured.

Copyright © Huawei Technologies Co., Ltd.

2. After the border router receives a packet not destined for the router from the IPv4 network, the router appends an IPv6 header to the IPv4 packet and encapsulates the IPv4 packet as an IPv6 packet.
3. On the IPv6 network, the encapsulated packet is transmitted to the remote border router.
4. The remote border router decapsulates the packet, removes the IPv6 header, and sends the decapsulated IPv4 packet to the IPv4 network.

5.3 References

The following table lists the references of this document.

Document	Description	Remarks
RFC1887	An Architecture for IPv6 Unicast Address Allocation	-
RFC1970	Neighbor Discovery for IP Version 6 (IPv6)	-
RFC1981	Path MTU Discovery for IP version 6	-
RFC2375	IPv6 Multicast Address Assignments	-
RFC2147	TCP and UDP over IPv6 Jumbograms	-
RFC2460	Internet Protocol, Version 6 (IPv6) Specification	-
RFC2461	Neighbor Discovery for IP Version 6 (IPv6)	-
RFC2462	IPv6 Stateless Address Auto configuration	-
RFC2463	Internet Control Message Protocol for the Internet Protocol Version 6 Specification	-
RFC2464	Transmission of IPv6 Packets over Ethernet Networks	-
RFC2472	IP Version 6 over PPP	-
RFC2473	Generic Packet Tunneling in IPv6 Specification	-
RFC2529	Transmission of IPv6 over IPv4 Domains without Explicit Tunnels	-
RFC2711	IPv6 Router Alert Option	-
RFC2893	Transition Mechanisms for IPv6 Hosts and Routers	-
RFC3056	Connection of IPv6 Domains via IPv4 Clouds	-

Copyright © Huawei Technologies Co., Ltd.

Document	Description	Remarks
RFC3068	An Anycast Prefix for 6to4 Relay Routers	-
RFC3484	Default Address Selection for Internet Protocol version 6 (IPv6)	-
RFC3493	Basic Socket Interface Extensions for IPv6	-
RFC3513	IP Version 6 Addressing Architecture	-
RFC3542	Advanced Sockets API for IPv6	-
RFC3587	An Aggregatable Global Unicast Address Format	-
RFC3879	Deprecating Site Local Addresses	-
RFC4007	IPv6 Scoped Address Architecture	-
RFC4193	Unique Local IPv6 Unicast Addresses	Currently, the part relating to DNS cannot be implemented.
RFC4213	Basic Transition Mechanisms for IPv6 Hosts and Routers	-
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)	Currently, the part relating to ND proxy and ND security cannot be implemented.
RFC4862	IPv6 Stateless Address Auto configuration	Currently, RFC4862 cannot be implemented on hosts.
RFC5095	Deprecation of Type 0 Routing Headers in IPv6	-

Copyright © Huawei Technologies Co., Ltd.

6 DNS

About This Chapter

6.1 Introduction to DNS

6.2 Principles

6.3 Applications

6.4 References

Copyright © Huawei Technologies Co., Ltd.

6.1 Introduction to DNS

Definition

Domain Name System (DNS) is a distributed database used in TCP and IP applications and completes resolution between IP addresses and domain names.

Purpose

Each host on the network is identified by an IP address. To access a host, a user must obtain the host IP address first. It is difficult for users to remember IP addresses of hosts. Therefore, host names in the format of strings are designed. Each host name maps an IP address. In this way, users can use the simple and meaningful domain names instead of the complicated IP addresses to access hosts.

6.2 Principles

6.2.1 Working Principle of DNS

Domain name resolution is classified into dynamic resolution and static resolution that complement each other. During domain name resolution, static resolution is preferentially used. If static resolution fails, dynamic resolution is used. To improve the domain name resolution efficiency, you are advised to add commonly used domain names to a static domain name resolution table.

Static DNS

A static domain name resolution table is manually set up, describing the mappings between domain names and IP addresses. Some common domain names are added to the table. To obtain the IP address by resolving a domain name,domain names are resolved based on the static domain name resolution table. In this manner, the efficiency of domain name resolution is improved.

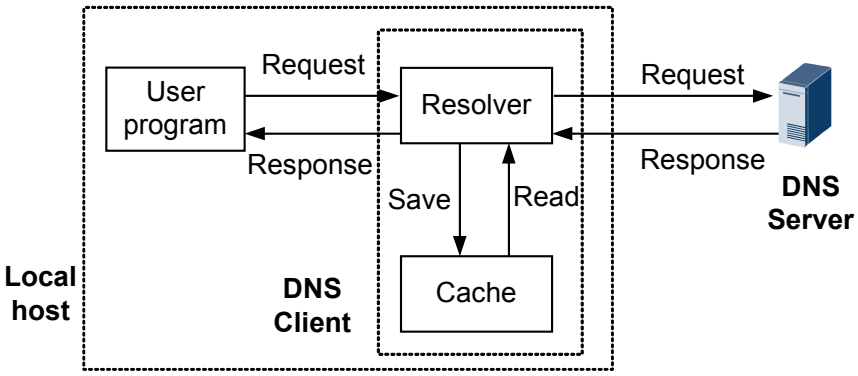
Dynamic DNS

User programs, such as ping and tracert, access the DNS server using the resolver of the DNS client.

Figure 6-1 shows the relationship between user programs, the resolver, the DNS server, and the cache on the resolver.

Copyright © Huawei Technologies Co., Ltd.

Figure 6-1 Dynamic DNS



The DNS client, consisting of the resolver and the cache, is used to accept and respond to the DNS queries from user programs. Generally, user programs(ping,Tracert), the cache, and the resolver are on the same host; whereas the DNS server is on another host.

Working Process of the Dynamic DNS

1. When a user accesses some applications by domain name, the user program sends a request to the resolver on the DNS client.
2. After receiving the request, the resolver searches the local domain name cache.
   - If the domain name matches an entry in the local cache, the resolver sends the corresponding IP address to the user program.
   - If the domain name matches no entry in the local cache, the resolver sends a query message to the DNS server.
3. When receiving the query message, the DNS server first checks whether the domain name to be resolved is in an authorized sub-domain. Then, the DNS server sends a response packet according to the check result.
   - If the domain name is in an authorized sub-domain, the DNS server searches for the corresponding IP address in the local database.
   - If the domain name is out of authorized sub-domains, the DNS server sends a query message to a higher-level DNS server. This process continues until the DNS server finds the corresponding IP address or detects that the corresponding IP address of the domain name does not exist. Then the DNS server returns a result to the DNS client.
4. After receiving the response packet from the DNS server, the DNS client sends the resolution result to the user program.

   Mappings between domain names and IP addresses are stored in the dynamic domain name cache. When resolving a domain name that is stored in the cache, the DNS client obtains the corresponding IP address from the cache directly and does not send a query message to the DNS server. Mappings stored in the cache will be deleted when the aging time expires to ensure that the latest mappings can be obtained from the DNS server. The aging time is set by the DNS server. The DNS client obtains the aging time from protocol packets.

Domain Name Suffix List

Dynamic domain name resolution supports the domain name suffix list. Users can preset domain name suffixes. Users only need to enter partial content of a domain name, and the system adds

Copyright © Huawei Technologies Co., Ltd.

a suffix to the domain name for resolution. For example, a user has set the domain name suffix com in the suffix list. To visit huawei.com, the user only needs to enter **huawei**. The system adds the suffix com to the domain name.

When the domain name suffix list is used, the resolution modes vary according to domain names entered by users.

- If a user enters a domain name without a dot (.), for example, **huawei**, the system identifies it as a host name and adds a suffix to the domain name for resolution. If the resolution fails, the system resolves the entered domain name.
- If a user enters a domain name with a dot (.) in the middle, for example, **www.huawei**, the system resolves the domain name. If the resolution fails, the system adds a suffix to the domain name for resolution.
- If a user enters a domain name with a dot (.) at the end, for example, **huawei.com.**, the system resolves only the entered domain name directly and sends a response packet regardless of whether the domain name is resolved correctly. The system does not add a preset suffix to the entered domain name for resolution. Therefore, the dot (.) at the end of the domain name is called query terminator. A domain name with a query terminator is an absolute domain name or a full qualified domain name (FQDN).

Domain Name Resolution Modes

Dynamic domain name resolution requires a special DNS server. This server provides mappings between domain names and IP addresses, and processes DNS client's request for domain name resolution.

After receiving a resolution request, the DNS server checks whether the domain name belongs to its authorized sub-domain. If so, the server searches the database and translates the domain name into an IP address, and sends the IP address to the DNS client. If the server fails to resolve the domain name, it performs the next operation according to the following resolution modes specified in the query packet:

- Recursive resolution

  The DNS server asks another server that can resolve the domain name for help, and then sends the resolved IP address to the DNS client.
- Iteration resolution

  The DNS server specifies another DNS server in the response packet for the DNS client to contact. Then, the DNS client sends a resolution request to the specified server.

Query Type

Class-A query is a common type of query, which is used to obtain the IP address corresponding to a specified domain name. For example, when you ping or tracert a domain name, the ping or tracert, as a user program, sends a query to the DNS client for the IP address corresponding to the domain name. If the corresponding IP address does not exist on the DNS client, the DNS client sends a Class-A query to the DNS server to obtain the corresponding IP address.

6.2.2 Working Principle of DNS Proxy or Relay

DNS proxy or relay is used to forward DNS request and reply packets between the DNS client and DNS server. The DNS client sends DNS request packets to the DNS proxy or relay. The DNS proxy or relay forwards request packets to the DNS server and sends reply packets to the

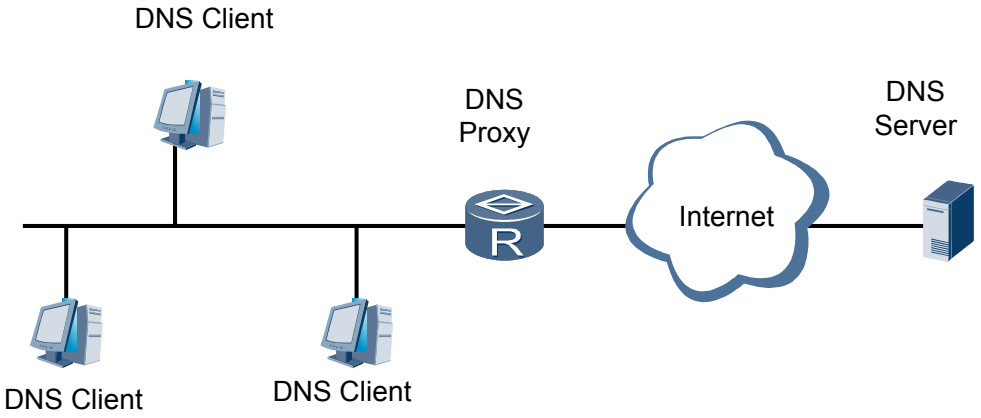
Copyright © Huawei Technologies Co., Ltd.

DNS client. After DNS proxy or relay is enabled, if the IP address of the DNS server changes, you only need to change the configuration on the DNS proxy or relay.

DNS relay is similar to DNS proxy. The difference is that the DNS proxy searches for DNS entries saved in the local domain name cache after receiving DNS query messages from DNS clients. The DNS relay, however, directly forwards DNS query messages to the DNS server, reducing the cache usage.

The application environments of DNS replay and DNS proxy are similar. **Figure 6-2** shows the typical networking of DNS proxy.

Figure 6-2 Typical networking of DNS proxy



The working process of DNS proxy is as follows:

1. The DNS client sends a request packet to the DNS proxy. The DNS proxy IP address is the destination address of the request packet.
2. After receiving the request packet, the DNS proxy searches for DNS entries saved in the local domain name resolution tables. If mapping information exists, the DNS proxy sends a reply packet carrying the resolution result to the DNS client.
3. If no mapping information exists, the DNS proxy sends the request packet to the DNS server for resolution.
4. After receiving the reply packet from the DNS server, the DNS proxy records the resolution result and forwards the reply packet to the DNS client.

Only when the IP address of the DNS server and the route to the DNS server exist on the DNS proxy, the DNS proxy sends domain name resolution requests to the DNS server. Otherwise, the DNS proxy neither sends any domain name resolution request to the DNS server nor replies any request from the DNS client.

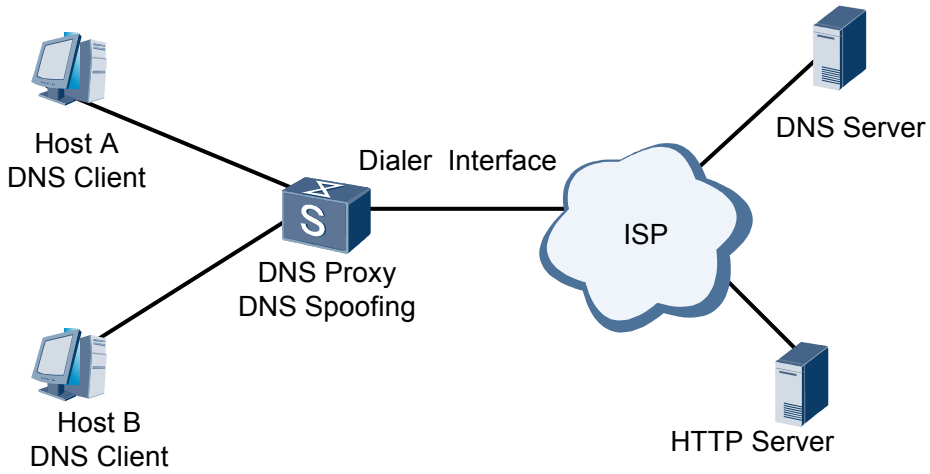
6.2.3 Working Principle of DNS Spoofing

When the DNS server IP address is not configured or the route to the DNS server does not exist on the DNS proxy or relay that is enabled with DNS spoofing, the DNS proxy or relay sends a spoofing IP address as the domain name resolution result to any DNS client that sends a DNS query message.

DNS spoofing is applied to a dial-up network, as shown in **Figure 6-3**.

Copyright © Huawei Technologies Co., Ltd.

Figure 6-3 DNS spoofing application scenario



As shown in **Figure 6-3**, the device functions as the DNS proxy and connects to the network using the dial-up interface. The dial-up interface is triggered to set up a connection only when data packets are forwarded by the dial-up interface. When the device functions as the DNS proxy, hosts A and B consider the device as the DNS server. When the dial-up connection is set up, the device obtains the DNS server IP address using DHCP.

When receiving a DNS query message from a DNS client, the device enabled with DNS spoofing sends a DNS query message to the DNS server when no matching entry is found. If the dial-up connection is not set up, the device cannot obtain the DNS server IP address. The device does not send a DNS query message to the DNS server or respond to the request from the DNS client. The domain name resolution fails. No data packet traffic triggers the dial-up interface to set up a connection.

DNS spoofing enables the device to send a spoofing IP address to the DNS client that sends a DNS query message regardless of whether the DNS server IP address is configured or the route to the DNS server exists on the device. Data packets sent by the DNS client triggers the dial-up interface to set up a connection.

As shown in **Figure 6-3**, a DNS client wants to access the HTTP server. The process is described as follows:

1. A DNS client sends a DNS query message to the DNS proxy for resolving the HTTP server domain name to an IP address.
2. After receiving the DNS query message, the DNS proxy cannot send the correct IP address to the DNS client because no matching entry is found locally, no dial-up connection is set up, and the DNS server IP address is not obtained. The DNS proxy sends the spoofing IP address as the resolution result to the DNS client. The aging time of a DNS resolution response message is 0. A reachable route between the DNS client and the IP address in the response message must exist. The outbound interface of the route is the dial-up interface.
3. After receiving the response message, the host sends an HTTP request to the IP address in the response message.
4. The DNS proxy forwards the HTTP request using the dial-up interface. The traffic triggers the dial-up interface to set up a connection with the DNS server. Then the DNS proxy obtains the DNS server IP address using DHCP.

Copyright © Huawei Technologies Co., Ltd.

5. After the DNS resolution response message is aged, the DNS client sends a DNS query message again.
6. The DNS proxy sends the correct IP address to the DNS client.
7. After obtaining the correct HTTP server IP address, the DNS client can access the HTTP server.

6.2.4 Working Principle of DDNS

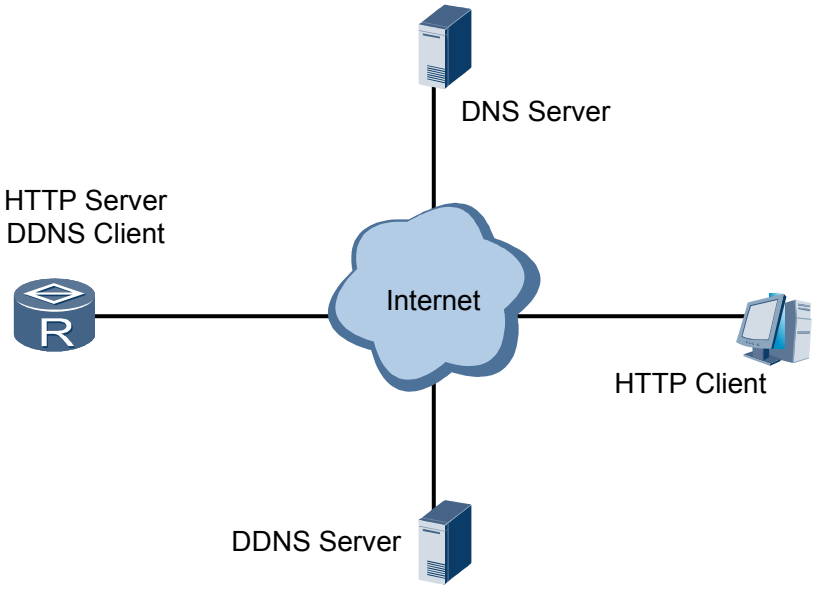
DDNS Overview

Dynamic domain name system (DDNS) resolves domain names into IP addresses so that you can access network nodes using domain names. DNS provides static mappings between domain names and IP addresses. When IP addresses of nodes change, DNS cannot dynamically update mappings. If a user uses the original domain name to access the node, the user will fail to access the node because the IP address mapping the domain name is incorrect. The Dynamic Domain Name System (DDNS) updates mappings between domain names and the IP addresses on the DNS server to ensure that the IP address can be resolved correctly.

DDNS Working Mode

DDNS works in client/server mode. **Figure 6-4** shows the typical networking of DDNS.

Figure 6-4 Typical networking of DDNS



As shown in **Figure 6-4**, DDNS works in client/server mode.

- When an IP address changes, the DDNS updates the mapping between the domain name and IP address on the DNS server. Internet users use domain names to access servers that provide application-layer services, such as HTTP and FTP servers. When the IP address of a server changes, the server functions as a DDNS client and sends a request for updating

Copyright © Huawei Technologies Co., Ltd.

the mapping between the domain name and the IP address to the DDNS server. This ensures that Internet users can still access a server when the server IP address changes.

- The DDNS server instructs the DNS server to dynamically update the mapping between the domain name and the IP address on the DNS server to ensure that the IP address can be resolved correctly and Internet users can access the DDNS client using the domain name.

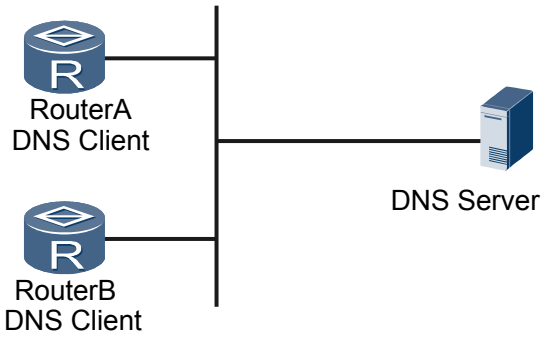
No unified standard is defined for the DDNS update process. DDNS update processes are different on different DDNS servers. DDNS servers provided at www.oray.cn, www.3322.org, and www.dyndns.com.

6.3 Applications

6.3.1 DNS Client Application

Figure 6-5 shows typical networking of a DNS client.

Figure 6-5 Typical networking of a DNS client



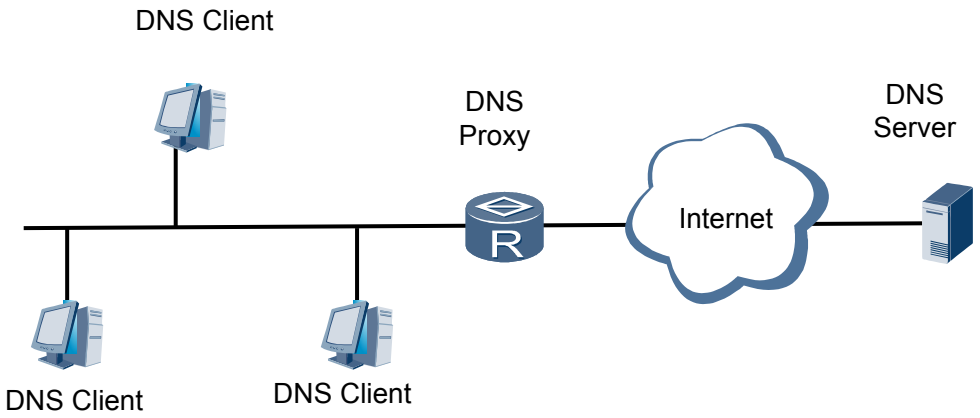
As shown in Figure 6-5, The device functions as a DNS client and can dynamically obtain the corresponding IP address of a domain name from a DNS server. This facilitates configurations and centralized management.

6.3.2 DNS Proxy Application

Figure 6-6 shows the typical networking of DNS proxy.

Copyright © Huawei Technologies Co., Ltd.

Figure 6-6 Typical networking of DNS proxy



The device functions as an egress router and is configured with DNS proxy in an enterprise. The device can forward DNS request and reply packets between DNS clients in the enterprise and DNS servers out of the enterprise. When the IP address of a DNS server changes, you only need to change the configuration on the DNS proxy,this will be beneficial to Network management.

6.4 References

The following table lists the references of this document:

Document	Description	Remarks
RFC1034	DOMAIN NAMES - CONCEPTS AND FACILITIES	-
RFC1035	DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION	-

Copyright © Huawei Technologies Co., Ltd.

7 NAT

About This Chapter

7.1 Introduction to NAT

7.2 Principles

7.3 Applications

7.4 References

Copyright © Huawei Technologies Co., Ltd.

7.1 Introduction to NAT

Definition

Network Address Translation (NAT) translates the IP address in an IP datagram header to another IP address.

Purpose

The rapid development of the Internet brings an increasing number of network applications. Exhaustion of IPv4 addresses has become a bottleneck for the network development. IPv6 can solve the problem of IPv4 address shortage, but numerous network devices and applications are based on IPv4. Major transitional technologies such as classless inter-domain routing (CIDR) and private network addresses are used before the wide use of IPv6 addresses. NAT enables users on private networks to access public networks. When a host on a private network accesses a public network, NAT translates the host's private IP address to a public IP address. Multiple hosts on a private network can share one public IP address. This implements network communication while saving public IP addresses. For the classification of private IP addresses, see **1.2.2 IPv4 Address**.

Benefits

As a transitional plan, NAT enables address reuse to meet the demand for IP addresses, therefore alleviating the IPv4 address shortage. In addition to solving the problem of IP address shortage, NAT provides the following advantages:

- Protects private networks against external attacks, greatly improving network security.
- Enables mutual access between hosts on private networks and public networks.

7.2 Principles

7.2.1 Overview

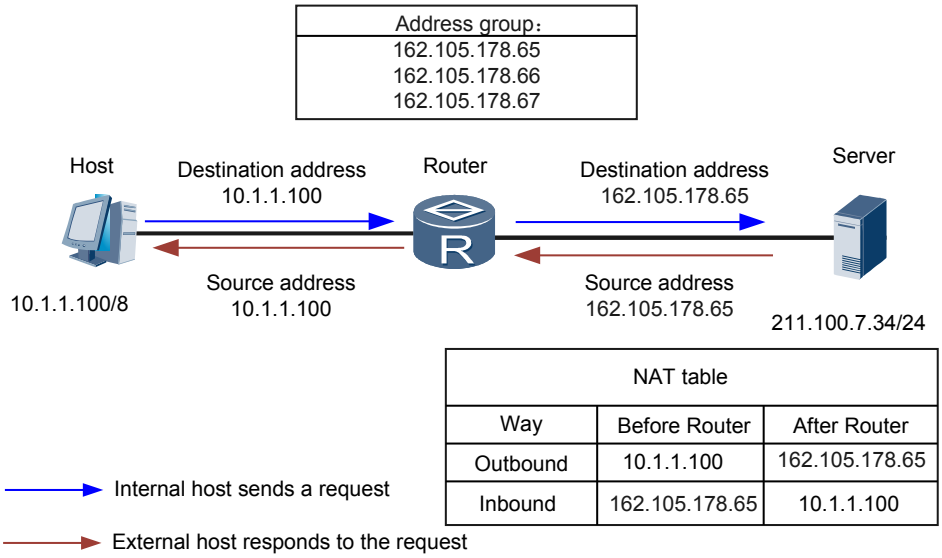
NAT translates the IP address in an IP datagram header to another IP address, allowing users on private networks to access public networks. Basic NAT implements one-to-one translation between one private IP address and one public IP address, whereas Network Address and Port Translation (NAPT) implements one-to-many translation between one public IP address and multiple private IP addresses.

Basic NAT

Basic NAT implements one-to-one IP address translation. In this mode, only the IP address is translated, whereas the TCP/UDP port number remains unchanged. Basic NAT cannot translate multiple private IP addresses to the same public IP address.

Copyright © Huawei Technologies Co., Ltd.

Figure 7-1 Networking diagram for basic NAT



As shown in **Figure 7-1**, the basic NAT process is as follows:

- The router receives a request packet sent from the host on the private network for accessing the server on the public network. The source IP address of the packet is 10.1.1.100.
- The router selects an idle public IP address (162.105.178.65) from the IP address pool, and sets up forward and reverse NAT entries that specify the mapping between the source IP address of the packet and the public IP address. The router translates the packet's source IP address to the public IP address based on the forward NAT entry, and sends the packet to the server on the public network. After the translation, the packet's source IP address is 162.105.178.65, and its destination IP address is 211.100.7.34.
- After receiving a response packet from the server on the public network, the router queries the reverse NAT entry based on the packet's destination IP address. The router translates the packet's destination IP address to the private IP address of the host on the private network based on the reverse NAT entry, and sends the packet to the host. After the translation, the packet's source IP address is 162.105.178.65, and its destination IP address is 10.1.1.100.

NOTE

Basic NAT cannot solve the problem of public IP address shortage because it cannot implement address reuse. Therefore, basic NAT is seldom used in practice.

The number of public IP addresses owned by the NAT server is far less than the number of hosts on private networks because not all the hosts on private networks access public networks at the same time. The number of public IP addresses needs to be determined based on the number of hosts on private networks that access public networks during peak hours.

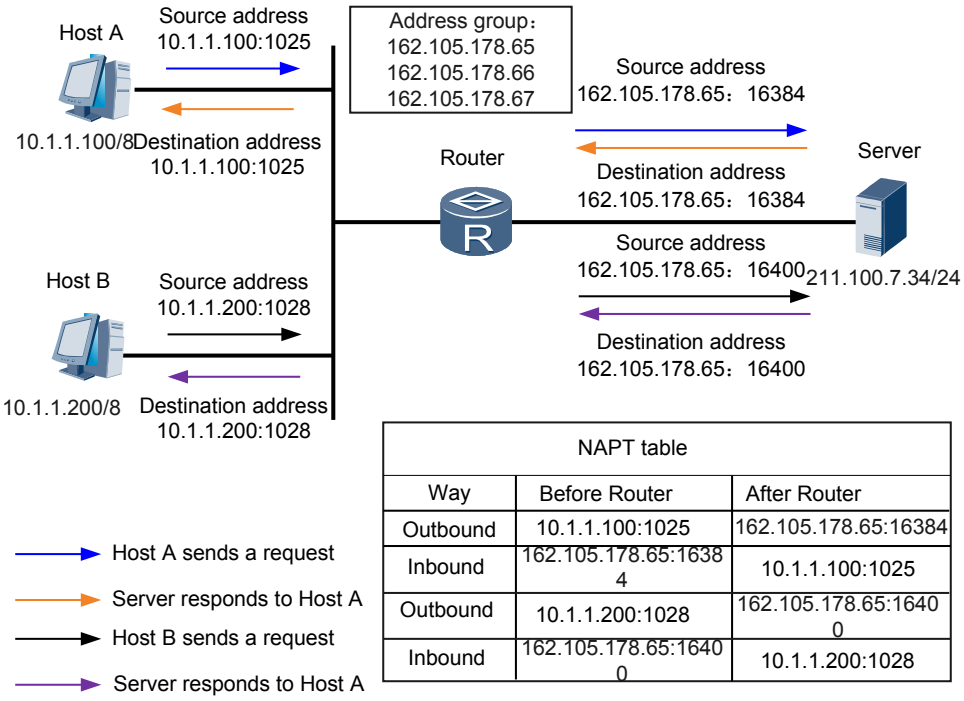
NAPT

In addition to one-to-one address translation, NAPT allows multiple private IP addresses to be mapped to the same public IP address. It is also called many-to-one address translation or address reuse.

NAPT translates the IP address and port number of a packet so that multiple users on a private network can use the same public IP address to access the public network.

Copyright © Huawei Technologies Co., Ltd.

Figure 7-2 Networking diagram for NAPT



NAPT table

Way	Before Router	After Router
Outbound	10.1.1.100:1025	162.105.178.65:16384
Inbound	162.105.178.65:16384	10.1.1.100:1025
Outbound	10.1.1.200:1028	162.105.178.65:16400
Inbound	162.105.178.65:16400	10.1.1.200:1028

As shown in **Figure 7-2**, the NAPT process is as follows:

- The router receives a request packet sent from the host on the private network for accessing the server on the public network. The packet's source IP address is 10.1.1.100, and its port number is 1025.
- The router selects an idle public IP address and an idle port number from the IP address pool, and sets up forward and reverse NAPT entries that specify the mapping between the source IP address and port number of the packet and the public IP address and port number. The router translates the packet's source IP address and port number to the public IP address and port number based on the forward NAPT entry, and sends the packet to the server on the public network. After the translation, the packet's source IP address is 162.105.178.65, and its port number is 16384.
- After receiving a response packet from the server on the public network, the router queries the reverse NAPT entry based on the packet's destination IP address and port number. The router translates the packet's destination IP address and port number to the private IP address and port number of the host on the private network based on the reverse NAPT entry, and sends the packet to the host. After the translation, the packet's destination IP address is 10.1.1.100, and its destination port number is 1025.

7.2.2 NAT Implementation

Basic NAT and NAPT translate private IP addresses to public IP addresses by using NAT devices. Basic NAT implements one-to-one address translation, and NAPT implements many-to-one address translation. On existing networks, NAT is implemented based on the principles of basic NAT and NAPT. NAT implements multiple functions such as Easy IP, NAT address pool, NAT server, and static NAT/NAPT.

NAT address pool and Easy IP are implemented in similar ways. This section describes only Easy IP. For the implementation of NAT address pool, see **7.2.1 Overview**.

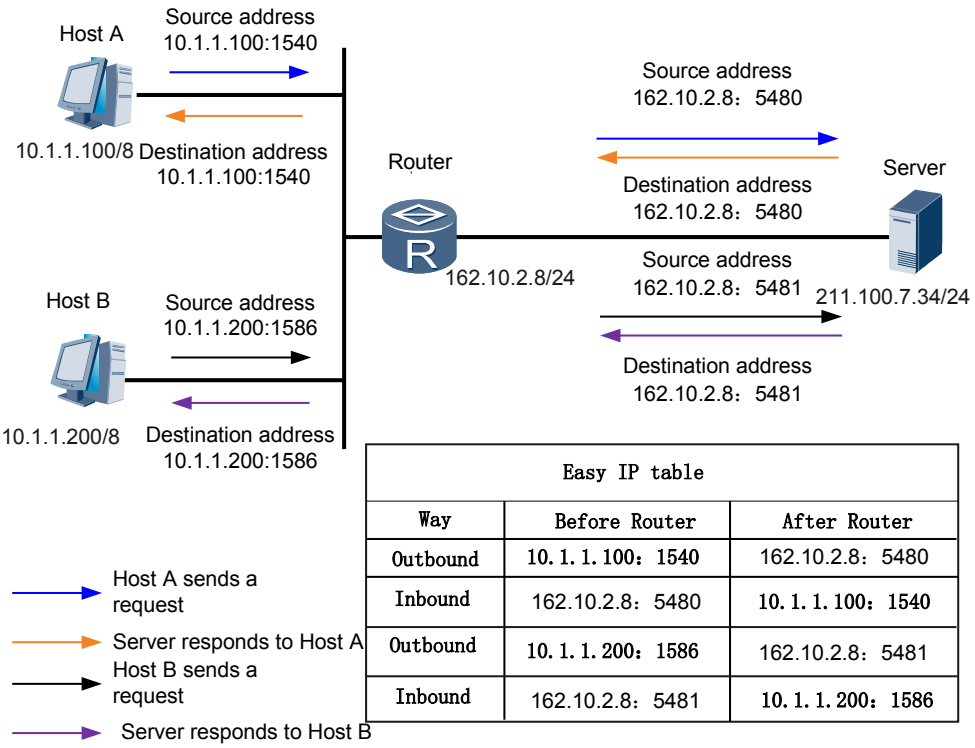
Copyright © Huawei Technologies Co., Ltd.

Easy IP

Easy IP uses access control lists (ACLs) to control the private IP addresses that can be translated.

Easy IP is applied to the scenario where hosts on small-scale LANs access the Internet. Small-scale LANs are usually deployed at small- and medium-sized cybercafes or small-sized offices where only a few internal hosts are used and the outbound interface obtains a temporary public IP address through dial-up. The temporary public IP address is used by the internal hosts to access the Internet. Easy IP allows the hosts to access the Internet using this temporary public address.

Figure 7-3 Networking diagram for Easy IP



Easy IP table

Way	Before Router	After Router
Outbound	10.1.1.100：1540	162.10.2.8：5480
Inbound	162.10.2.8：5480	10.1.1.100：1540
Outbound	10.1.1.200：1586	162.10.2.8：5481
Inbound	162.10.2.8：5481	10.1.1.200：1586

As shown in **Figure 7-3**, the Easy IP process is as follows:

1. The router receives a request packet sent from the host on the private network for accessing the server on the public network. The packet's source IP address is 10.1.1.100, and its port number is 1540.
2. The router sets up forward and reverse Easy IP entries that specify the mapping between the source IP address and port number of the packet and the public IP address and port number of the port connected to the public network. The router translates the source IP address and port number of the packet to the public IP address and port number based on the forward Easy IP entry, and sends the packet to the server on the public network. After the translation, the packet's source IP address is 162.10.2.8, and its port number is 5480.
3. After receiving a response packet from the server on the public network, the router queries the reverse Easy IP entry based on the packet's destination IP address and port number. The router translates the packet's destination IP address and port number to the private IP address and port number of the host on the private network based on the reverse Easy IP entry, and

Copyright © Huawei Technologies Co., Ltd.

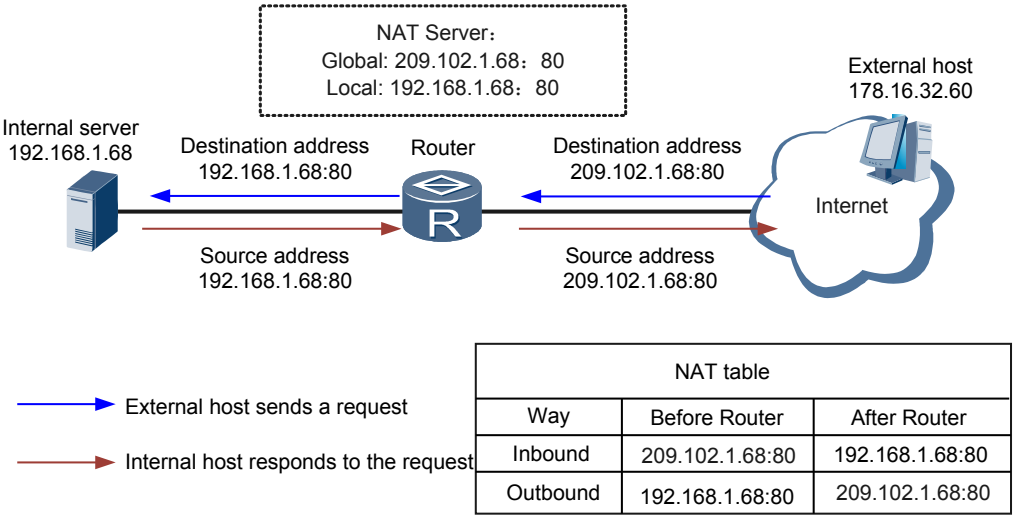
sends the packet to the host. After the translation, the packet's destination IP address is 10.1.1.100, and its port number is 1540.

NAT Server

NAT can shield hosts on private networks from public network users. When a private network needs to provide services such as web and FTP services for public network users, servers on the private network must be accessible to public network users at any time.

The NAT server can address the preceding problem by translating the public IP address and port number to the private IP address and port number based on the preset mapping.

Figure 7-4 Networking diagram for NAT server implementation



NAT table		
Way	Before Router	After Router
Inbound	209.102.1.68:80	192.168.1.68:80
Outbound	192.168.1.68:80	209.102.1.68:80

As shown in **Figure 7-4**, the address translation process of the NAT server is as follows:

1. Address translation entries of the NAT server are configured on the router.
2. The router receives an access request sent from a host on the public network. The router queries the address translation entry based on the packet's destination IP address and port number. The router translates the packet's destination IP address and port number to the private IP address and port number based on the address translation entry, and sends the packet to the server on the private network. The destination IP address of the packet sent by the host on the public network is 209.102.1.68, and its port number is 80. After the translation by the router, the destination IP address of the packet is 192.168.1.68, and its port number remains unchanged.
3. After receiving a response packet sent from the server on the private network, the router queries the address translation entry based on the packet's source IP address and port number. The router translates the packet's source IP address and port number to the public IP address and port number based on the address translation entry, and sends the packet to the host on the public network. The source of the response packet sent from the host on the private network is 192.168.1.68, and its port number is 80. After translation by the router, the source IP address of the packet is 209.102.1.68, and its port number remains unchanged.

Static NAT/NAPT

Static NAT indicates that a private IP address is statically bound to a public IP address when NAT is performed. Only this private IP address can be translated to this public IP address.

Copyright © Huawei Technologies Co., Ltd.

Static NAPT indicates that the combination of a private IP address, protocol number, and port number is statically bound to the combination of a public IP address, protocol number, and port number. Multiple private IP addresses can be translated to the same public IP address.

Static NAT and static NAPT can translate the IP address of a host in a specified range on the private network to an IP address within the specified public network segment. Static NAT or static NAPT translates only the network segment address, and host addresses remain unchanged. When a host on a private network accesses a public network, static NAT or static NAPT translates the IP address of the host to a public address if the IP address of the host is in the specified address range. When a host on a public network accesses a private network, static NAT or static NAPT translates the public IP address to a private IP address, which is in the specified address range. Then, the host on the public network can access the private network.

7.2.3 NAT ALG

NAT and NAPT can translate only IP addresses in IP datagram headers and port numbers in TCP/UDP headers. For some special protocols such as FTP, IP addresses or port numbers may be contained in the Data field of the protocol packets. Therefore, NAT cannot translate the IP addresses or port numbers.

For example, when an FTP server with a private IP address sets up a session with a host on the public network, the server may need to send its IP address to the host. NAT cannot translate this IP address because the IP address is carried in the Data field. When the host on the public network attempts to use the received private IP address, it finds that the FTP server is unreachable.

DNS, FTP, SIP, PPTP and RTSP support the ALG function.

Table 7-1 Fields translated by ALG in application protocol packets

Application Protocol	Field
DNS	IP and Port fields in a response packet
FTP	● IP and Port fields in the payload of a Port request packet ● IP and Port fields in the payload of a Passive response packet
SIP	● Request line ● From ● To ● Contact ● Via ● O ● Connection information field (indicating an IP address) and media description field (indicating a port) in the Message body
PPTP	There are two scenarios: ● PPTP client on the private network and PPTP server on the public network: Client-Call-ID field ● PPTP server on the private network and PPTP client on the public network: Server-Call-ID field

Copyright © Huawei Technologies Co., Ltd.

Application Protocol	Field
RTSP	Port field in a setup/reply OK packet

ALG Processing Mechanism

A good way to solve the NAT issue for these special protocols is to use the application level gateway (ALG) function. As a special translation agent for application protocols, the ALG interacts with the NAT device to establish states. It uses NAT state information to change the specific data in the Data field of IP datagrams and complete other necessary work, so that application protocols can run across private and public networks.

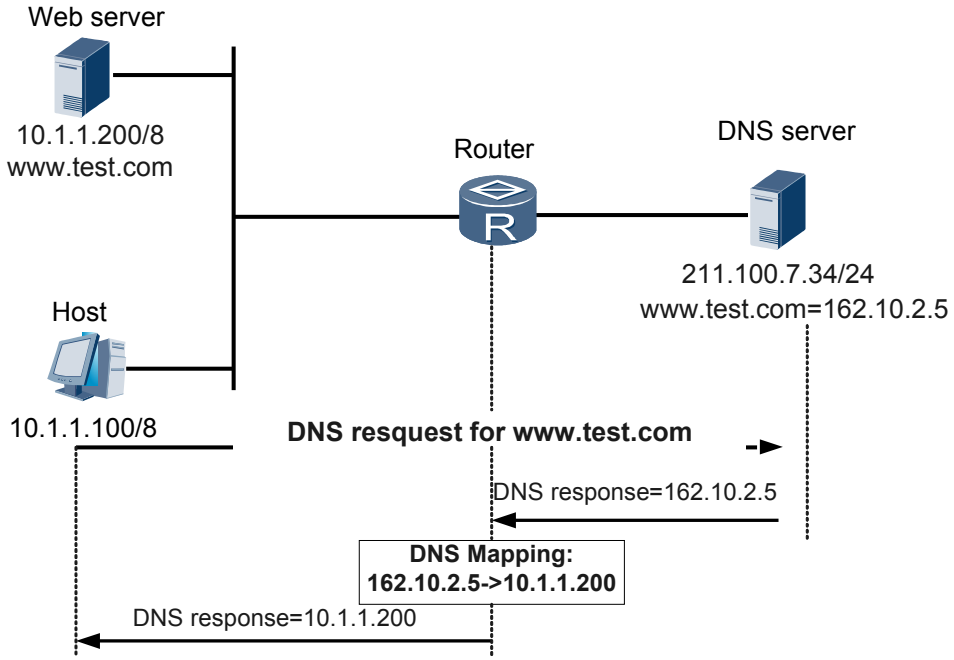
7.2.4 DNS Mapping

In practice, users on a private network need to access internal servers on the same private network using domain names, but the DNS server is located on a public network. Usually, a DNS response packet carries the public IP address of an internal server. If the NAT device does not replace the public IP address resolved by the DNS server with the private IP address of the internal server, users on the private network cannot access the internal server using the domain name.

DNS mapping can solve the problem by configuring a table that specifies the mapping between domain names, public IP addresses, public port numbers, and protocol types. In this manner, the mapping between domain names of servers on the private network and public network information is established.

Figure 7-5 describes the implementation of DNS mapping.

Figure 7-5 Networking diagram for DNS mapping



As shown in Figure 7-5, the host on the private network needs to access the web server using the domain name, and the router functions as a NAT server. After receiving a DNS response

Copyright © Huawei Technologies Co., Ltd.

packet, the router searches the DNS mapping table for the information about the web server based on the domain name carried in the response packet. Then, the router replaces the public IP address carried in the DNS response packet with the private IP address of the web server. In this manner, the DNS response packet received by the host carries the private IP address of the web server. Then, the host can access the web server using the domain name.

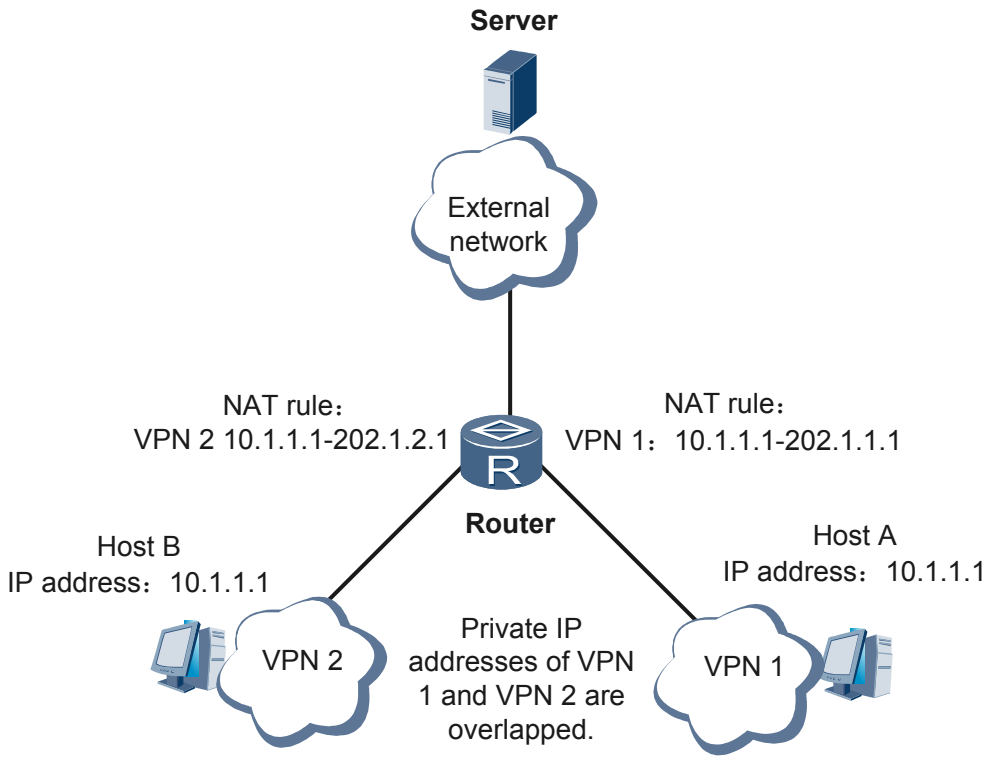
7.2.5 NAT Associated with VPNs

A NAT-enabled router allows hosts on private networks to access public networks, hosts in different virtual private networks (VPNs) on a private network to access a public network through the same outbound interface, and hosts with the same IP address in different VPNs to access a public network simultaneously. The NAT module of a router also supports NAT server associated with VPNs. It allows a host on a public network to access hosts in different VPNs on a private network, and a host on a public network to access hosts with the IP address in different VPNs on a private network.

Source NAT Associated with VPNs

Source NAT associated with VPNs allows hosts in different VPNs on a private network to access a public network using NAT. **Figure 7-6** shows the networking for NAT associated with VPNs.

Figure 7-6 Networking diagram for source NAT associated with VPNs



Source NAT associated with VPNs is implemented as follows:

1. The IP addresses of host A in VPN 1 and host B in VPN 2 are 10.1.1.1. Host A and host B want to access the same server on the public network.

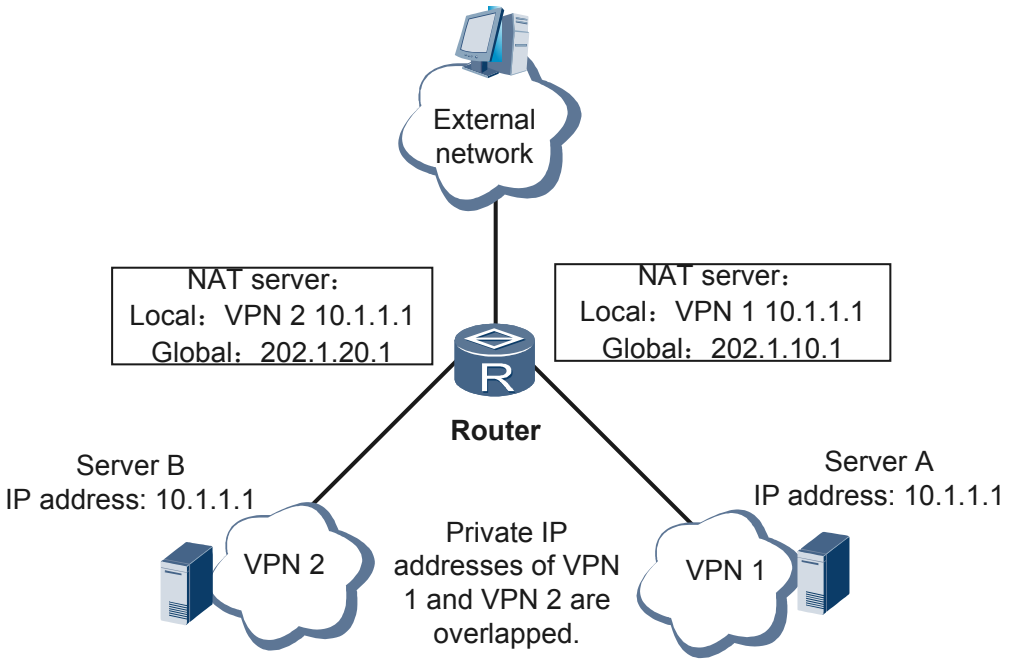
Copyright © Huawei Technologies Co., Ltd.

2. When a router functions as a NAT device, the router translates the source IP address of the packet sent from host A to 202.1.1.1 and the source IP address of the packet sent from host B to 202.1.2.1. In addition, the router records the VPN information about the hosts in the NAT translation table.
3. When the response packets sent from the server on the public network to host A and host B pass through the router:
   - The NAT module translates the destination IP address 202.1.1.1 of the packet sent to host A to 10.1.1.1 based on the NAT translation table, and then sends the packet to host A in VPN 1.
   - The NAT module translates the destination IP address 202.1.2.1 of the packet sent to host B to 10.1.1.1 based on the NAT translation table, and then sends the packet to host B in VPN 2.

NAT Server Associated with VPNs

NAT server associated with VPNs allows hosts on a public network to access servers in different VPNs on a private network using NAT. **Figure 7-7** shows the networking for the NAT server associated with VPNs.

Figure 7-7 Networking diagram for NAT server associated with VPNs



As shown in **Figure 7-7**, the IP addresses of server A in VPN 1 and server B in VPN 2 are 10.1.1.1. The public address of server A is 202.1.10.1 and that of server B is 202.1.20.1. Hosts on the public network can access server A using 202.1.10.1 and access server B using 202.1.20.1.

The NAT server associated with VPNs is implemented as follows:

1. A host on the public network sends a packet with the destination IP address as 202.1.10.1 to server A in VPN 1 and sends a packet with the destination IP address as 202.1.20.1 to server B in VPN 2.

Copyright © Huawei Technologies Co., Ltd.

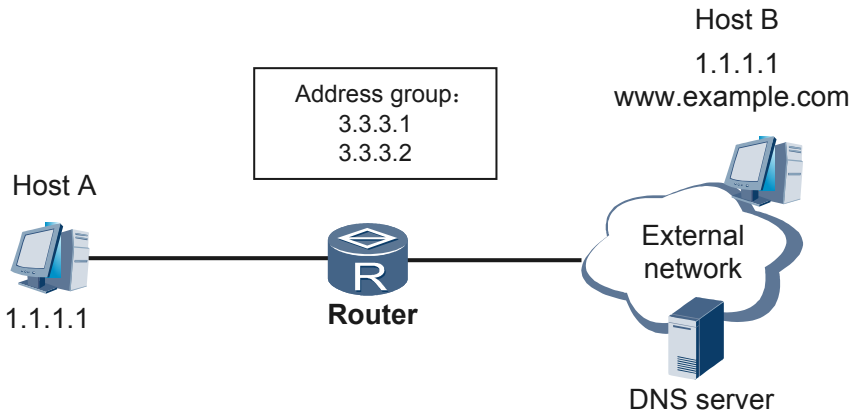
2. The router functions as the NAT server. Based on the packets' destination IP addresses and VPN information:
   - The router translates the destination address 202.1.10.1 to 10.1.1.1 and sends the packet to server A in VPN 1.
   - The router translates the destination address 202.1.20.1 to 10.1.1.1 and sends the packet to server B in VPN 2.

   In addition, the router records the VPN information in the NAT translation table.
3. When the response packets sent from server A and server B to the host on the public network pass through the router:
   - The NAT module translates the source IP address 10.1.1.1 of the packet sent from server A to 202.1.10.1 based on the NAT translation table, and sends the packet to the host on the public network.
   - The NAT module translates the source IP address 10.1.1.1 of the packet sent from server B to 202.1.20.1 based on the NAT translation table, and sends the packet to the host on the public network.

7.2.6 Twice NAT

Twice NAT refers to translation of both the source and destination IP addresses of a data packet. It is applied to the situation where a private IP address is the same as a public IP address.

Figure 7-8 Networking diagram for twice NAT



The process of twice NAT is described as follows:

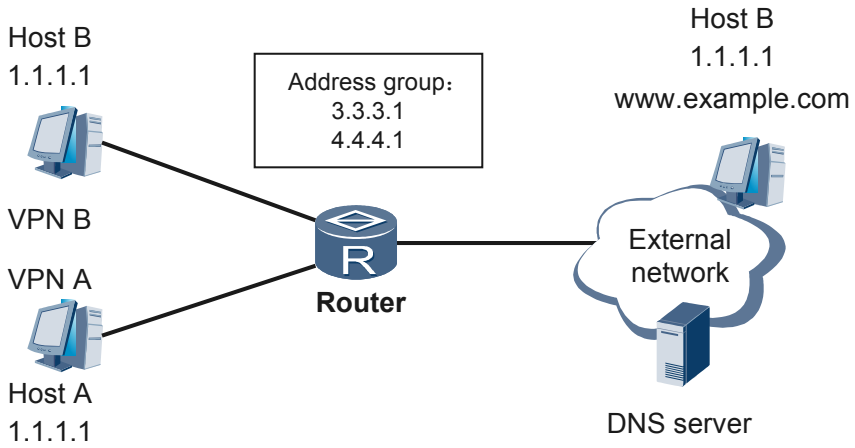
1. Host A with the IP address 1.1.1.1 on the private network wants to access host B with the same IP address on the public network. Host A sends a DNS request to the DNS server on the public network. The DNS server sends a response packet containing the IP address 1.1.1.1 of host B. When the response packet passes through the router, the router performs DNS ALG and translates host B's IP address 1.1.1.1 in the response packet to the unique temporary IP address 3.3.3.1. Then, the router forwards the response packet to Host A.
2. Host A sends a request packet with the destination IP address as the temporary IP address 3.3.3.1, for accessing host B. When the request packet passes through the router, the router detects that the destination IP address is the temporary IP address, and translates the destination IP address to host B's real IP address 1.1.1.1. Meanwhile, the router translates

Copyright © Huawei Technologies Co., Ltd.

the source IP address of the request packet to an address in the outbound NAT address pool using outbound NAT. Then, the router forwards the request packet to host B.

3. Host B sends host A a response packet with the destination IP address as the address in the outbound NAT address pool and the source IP address as the IP address of host B 1.1.1.1. When the response packet passes through the router, the router detects that the source IP address is the same as the real IP address of host A, and translates the source IP address to the temporary IP address 3.3.3.1 using NAT. Meanwhile, the router translates the destination IP address of the response packet to the private IP address 1.1.1.1 of host A. Then, the router forwards the response packet to host A.

Figure 7-9 Networking diagram for twice NAT when multiple VPNs are deployed on a private network



A private network may consist of multiple VPNs and hosts in the VPNs may have the same IP address. When configuring DNS ALG on a router, you need to add the VPN information that is used as the condition for mapping identical IP addresses of the hosts in the VPNs to IP addresses in the temporary address pool. **Figure 7-9** shows the networking for twice NAT when multiple VPNs are deployed on a private network. When multiple VPNs are deployed on a private network, the twice NAT process remains unchanged. The source IP address of host A in VPN A is translated to the temporary address 3.3.3.1, and the source IP address of host B in VPN B is translated to the temporary address 4.4.4.1.

7.2.7 NAT Filtering and NAT Mapping

NAT filtering allows an NAT device to filter the traffic from a public network to a private network. NAT mapping enables the IP addresses of a group of hosts on a private network to be mapped to the same public IP address using the NAT mapping table.

NAT Filtering

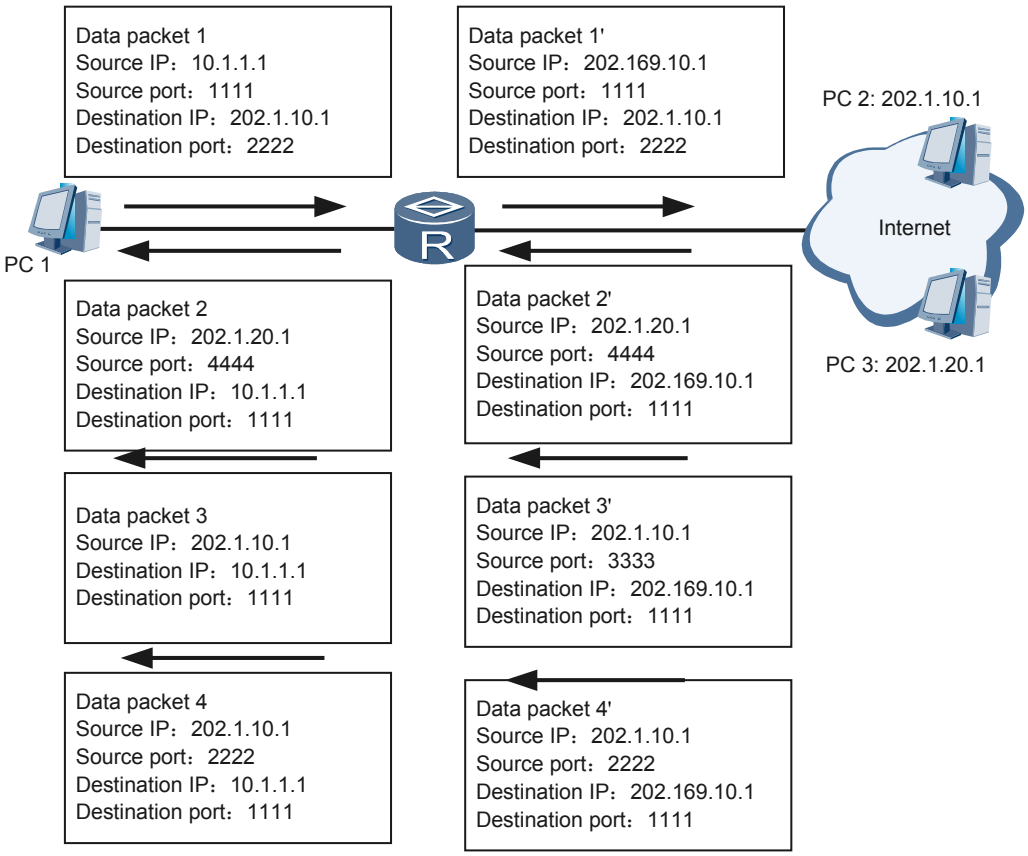
A NAT device filters the traffic from external network to internal network. NAT filtering includes the following modes:

- Endpoint-independent filtering
- Endpoint-dependent filtering
- Endpoint and port-dependent filtering

Copyright © Huawei Technologies Co., Ltd.

Figure 7-10 shows the NAT filtering applications.

Figure 7-10 NAT filtering applications



As shown in the preceding figure, PC-1 on the private network communicates with PC-2 and PC-3 on the public network using a NAT device. Datagram 1 is sent from PC-1 to PC-2. The source port number of the datagram is 1111 and the destination port number is 2222. The NAT device translates the source IP address to 202.169.10.1.

After PC-1 sends an access request to a PC on the public network, the PC on the public network transmits traffic to PC-1, and the NAT device filters the traffic destined for PC-1. Datagram 2, datagram 3, and datagram 4 are sent in three scenarios corresponding to the preceding three NAT filtering modes.

- Datagram 2 is sent from PC-3 to PC-1. The destination address of datagram 2 is different from that of datagram 1, and the destination port number is 1111. Datagram 2 can pass through the NAT device only when endpoint-independent filtering is used.
- Datagram 3 is sent from PC-2 to PC-1. The destination address of datagram 3 is the same as that of datagram 1, and the destination port number is 1111. The source port number of datagram 3 is 3333, which is different from that of datagram 1. Datagram 3 can pass through the NAT device only when endpoint-dependent filtering or endpoint-independent filtering is used.
- Datagram 4 is sent from PC-2 to PC-1. The destination address of datagram 4 is the same as that of datagram 1, and the destination port number is 1111. The source port number of datagram 4 is 2222, which is the same as that of datagram 1. In this case, endpoint and port-

Copyright © Huawei Technologies Co., Ltd.

dependent filtering is used, which is the default one. Datagram 4 can pass through the NAT device no matter whether a filtering mode is configured or no matter which filtering mode is configured.

A router supports the three NAT filtering modes.

NAT Mapping

After NAT mapping is enabled on a public network, it seems that all flows from a private network come from the same IP address because hosts on the private network share the same public IP address. When a host on the private network initiates a session request to a host on the public network, the NAT device searches the NAT translation table for the related session record. If the NAT device finds the session record, it translates the private IP address and port number and forwards the request. If the NAT device does not find the session record, it translates the private IP address and port number and meanwhile adds a session record to the NAT translation table. NAT mapping includes the following modes:

- Endpoint-independent mapping: The NAT uses the same IP address and port mapping for packets sent from the same private IP address and port to any public IP address and port.
- Endpoint-dependent mapping: The NAT uses the same port mapping for packets sent from the same private IP address and port to the same public IP address, regardless of the public port.
- Endpoint and port-dependent mapping: The NAT uses the same port mapping for packets sent from the same private IP address and port to the same public IP address and port if the mapping is still active.

A router supports endpoint-independent and endpoint and port-dependent mapping.

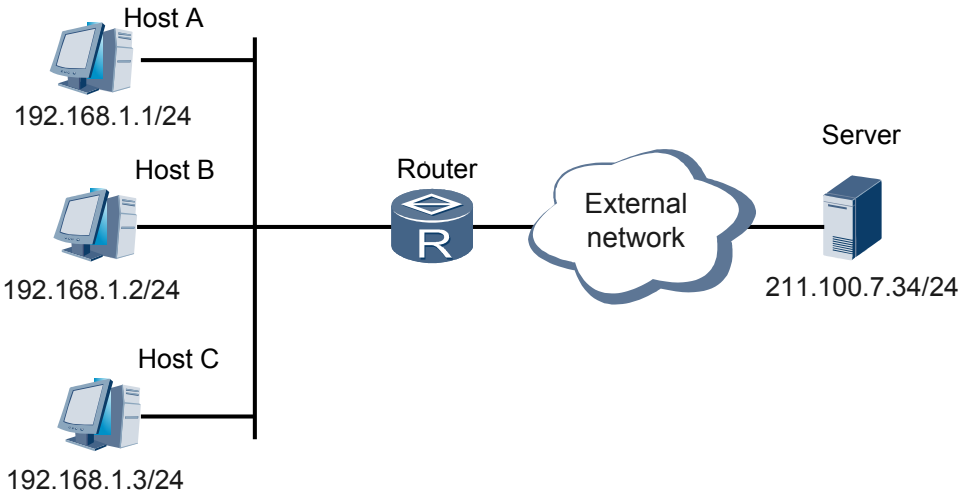
7.3 Applications

7.3.1 Private Network Hosts Accessing Public Network Servers

Private IP addresses are planned for hosts on private networks for communities, schools, and enterprises because public IP addresses are limited. In this case, the NAT technology can be used to implement access from hosts on the private networks to public networks. As shown in **Figure 7-11**, NAPT is configured on the router to enable the hosts on the private network to access the server on the public network.

Copyright © Huawei Technologies Co., Ltd.

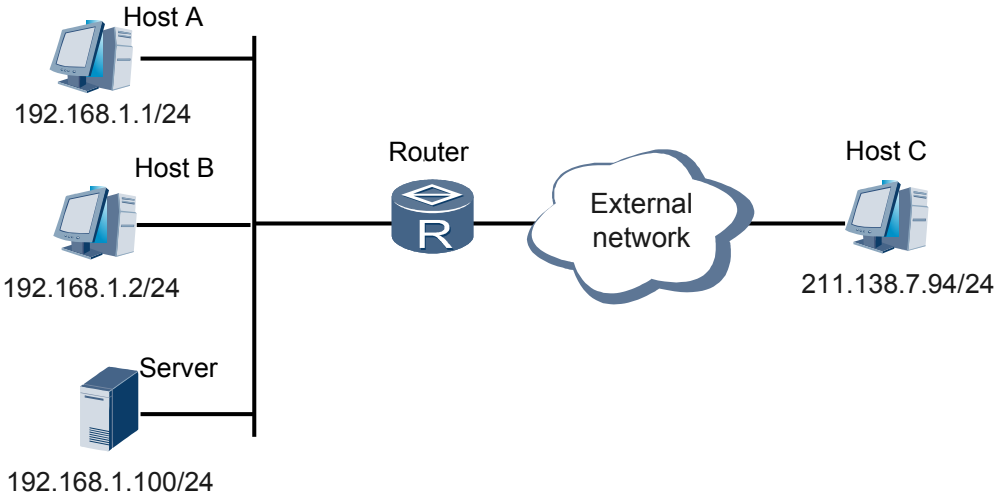
Figure 7-11 Networking diagram for private network hosts accessing public network servers



7.3.2 Public Network Hosts Accessing Private Network Servers

On private networks, some servers such as web servers and FTP servers need to provide services for public network users. NAT supports this application. As shown in **Figure 7-12**, the NAT server is configured. That is, mapping between the public IP address and port number and the private IP address and port number is defined. As a result, the host on the public network can access the server on the private network using the mapping.

Figure 7-12 Networking diagram for public network hosts accessing private network servers



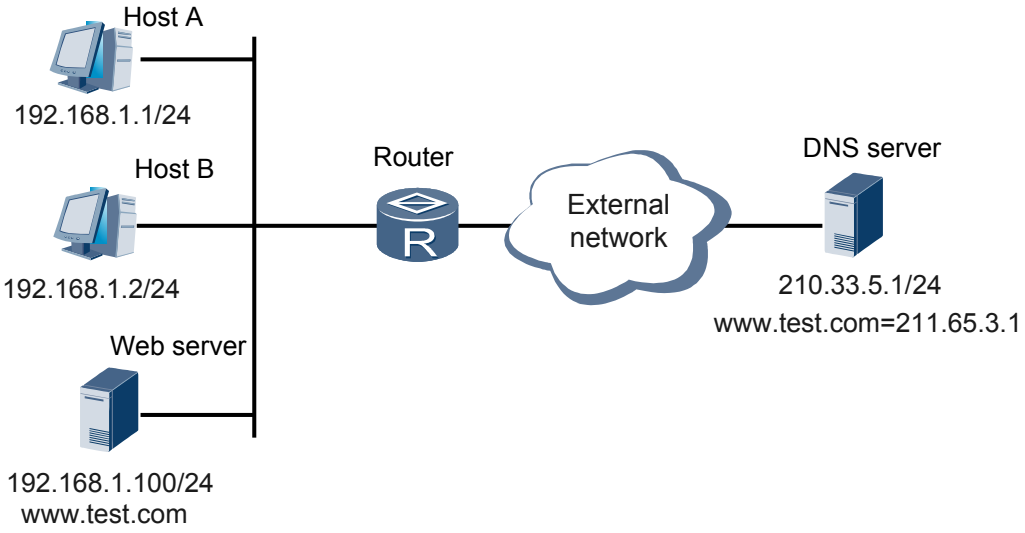
7.3.3 Private Network Hosts Accessing Private Network Servers Using the Domain Name

Hosts on a private network need to access a server on the same private network using the domain name. The DNS server, however, is located on a public network. You can configure DNS

Copyright © Huawei Technologies Co., Ltd.

mapping to allow the private network hosts to access the DNS server. As shown in **Figure 7-13**, a DNS mapping table is configured to define mapping between the domain name, public IP address, public IP address, and protocol type. The public IP address carried in the DNS response packet is replaced by the private IP address of the server on the private network. In this manner, hosts on the private network can access the server using the domain name.

Figure 7-13 Networking diagram for private network hosts accessing private network servers using the domain name

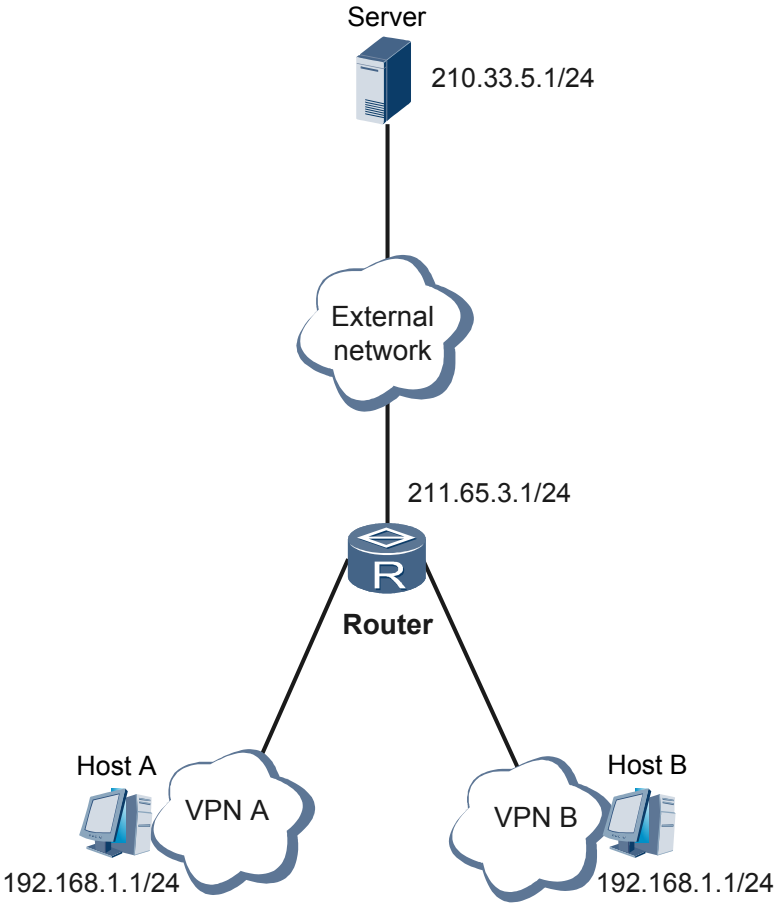


7.3.4 NAT Multi-instance

NAT multi-instance allows hosts that belong to different MPLS VPNs but have the same private IP address to access a public network through the same egress device simultaneously. As shown in **Figure 7-14**, host A and host B have the same private IP address, but they belong to different VPNs. NAT associated with VPNs is enabled to differentiate the hosts in different VPNs. In this manner, host A and host B can access the public network server simultaneously.

Copyright © Huawei Technologies Co., Ltd.

Figure 7-14 Networking diagram for NAT multi-instance



7.4 References

The following table lists the references of this document.

Document	Description
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)

Copyright © Huawei Technologies Co., Ltd.

Document	Description
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

Copyright © Huawei Technologies Co., Ltd.

8 IP Unicast Policy-based Routing

About This Chapter

8.1 Introduction to IP Unicast Policy-based Routing
8.2 Principles
8.3 Applications
8.4 References

Copyright © Huawei Technologies Co., Ltd.

8.1 Introduction to IP Unicast Policy-based Routing

Definition

Policy-based routing (PBR) is a mechanism that makes routing decisions based on user-defined policies. PBR includes local PBR, interface PBR, and smart policy routing (SPR).

NOTE

- The differences between PBR and routing policy are as follows:
  - PBR implements routing based on packets. It routes data packets based on user-defined policies instead of following the routes in the existing routing table.
  - Routing policies implement routing based on routing information. Routing policies are used to filter routes and set route attributes. You can change route attributes (including reachability) to change a route over which network traffic is transmitted.

  For details on routing policies, see Routing Policy in the *Feature Description - IP Routing*.
- IPv6 unicast PBR applies to IPv6 unicast packets. Its implementation is similar to that of IPv4 unicast PBR. Currently, IPv6 unicast PBR supports only interface PBR.

Purpose

Traditionally, devices searches routing tables for routes of packets based on their destination addresses and then forward the packets. Currently, more users require that devices route packets based on user-defined policies. IP unicast PBR allows network administrators to make user-defined policies to change packet routes based on source addresses, packet size, and link quality in addition to destination addresses.

Benefits

IP unicast PBR has the following advantages:

- Allows network administrators to make user-defined policies for routing packets, which improves flexibility of route selection.
- Allows different data flows to be forwarded on different links, which increases link usage.
- Uses cost-effective links to transmit service data without affecting service quality, which reduces the cost of enterprise data services.

8.2 Principles

8.2.1 Local PBR

Local PBR applies only to locally generated packets, such as ping packets.

Local PBR on a device can have multiple local PBR nodes. Each local PBR node has a priority. The device attempts to match locally generated packets with rules bound with local PBR nodes in descending order of priority.

Copyright © Huawei Technologies Co., Ltd.

Implementation

When sending locally generated packets, a device attempts to match the packets with rules bound with local PBR nodes in descending order of priority. Local PBR supports rules based on access control list (ACL) and packet length.

- If the device finds a matching local PBR node, it performs the following steps:
  1. Checks whether the priority of the packets has been set.
     - If so, the device applies the configured priority to the packets and performs step 2.
     - If not, the device performs step 2.
  2. Checks whether an outbound interface has been configured for local PBR.
     - If so, the device sends the packets through the outbound interface.
     - If not, the device performs step 3.
  3. Checks whether next hops have been configured for local PBR.

     NOTE

     Two next hops can be configured for load balancing.

     - If a next hop is configured for packet forwarding in PBR and the next hop is reachable, the device checks whether association between next hop and route is configured.
       - If association between next hop and route is configured, the device detects whether the configured IP address of the route associated with the next hop in PBR is reachable.
         - If the IP address is reachable, the configured next hop takes effect, and the device forwards packets to the next hop.
         - If the IP address is unreachable, the configured next hop does not take effect, and the device checks whether a backup next hop has been configured. If a backup next hop has been configured and is reachable, the device forwards packets to the backup next hop. If no backup next hop is configured or the configured backup next hop is unreachable, the device searches the routing table for a route according to the destination of packets. If no route is available, the device performs step 4.
       - If association between next hop and route is not configured, the device sends packets to the next hop.
     - If a next hop is configured in PBR but the next hop is unreachable, the device checks whether a backup next hop has been configured. If a backup next hop has been configured and is reachable, the device forwards packets to the backup next hop. If no backup next hop is configured or the configured backup next hop is unreachable, the device searches the routing table for a route according to the destination of packets. If no route is available, the device performs step 4.
     - If the next hop is not configured, the device searches the routing table for a route based on the destination addresses of the packets. If no route is available, the device performs step 4.
  4. Checks whether a default outbound interface has been configured for local PBR.
     - If so, the device sends the packets through the default outbound interface.
     - If not, the device performs step 5.
  5. Checks whether default next hops have been configured for local PBR.

Copyright © Huawei Technologies Co., Ltd.

- If so, the device sends the packets to the default next hops.
- If not, the device performs step 6.

6. Discards the packets and generates ICMP_UNREACH messages.

- If the device does not find a matching local PBR node, it searches the routing table for a route based on the destination addresses of the packets and then sends the packets.

8.2.2 Interface PBR

Interface PBR applies only to packets received from other devices, but not to locally generated packets such as local ping packets.

Implementation

Interface PBR is implemented based on the redirect action configured in a traffic behavior and takes effect only on the inbound packets. By default, a device forwards packets to the next hop found in the routing table. If interface PBR is configured, the device forwards packets to the next hop specified by interface PBR.

When the device forwards packets to the next hop specified by interface PBR, the device triggers ARP learning if it has no ARP entry corresponding to the IP address of the specified next hop. If the device cannot learn this ARP entry, it forwards packets to the next hop found in the routing table. If the device has this ARP entry, it forwards packets to the next hop specified by interface PBR.

8.2.3 Smart Policy Routing

Smart policy routing (SPR) allows devices to select routes based on link quality and requirements for link quality.

Background

As network service requirements vary widely and service data is stored in a centralized manner, network services increasingly depend on high link quality. Users stress more importance on service availability, service response speed, and service quality than network connectivity. Diversified service requirements pose a challenge to per-hop-based routing protocols. Devices that use per-hop-based routing protocols are unaware of link quality and service requirements, so they cannot deliver satisfying service experience. Even though routes are reachable, low link quality may cause a packet forwarding failure. SPR resolves this problem. SPR selects optimal links to forward packets based on link quality and service requirements. This mechanism prevents network black holes and flappings.

Service Classification

SPR classifies services based on the following attributes:

- Protocol types: IP, TCP, UDP, GRE, IGMP, IPINIP, OSPF, and ICMP
- Packet applications: DSCP, TOS, IP precedence, fragment, VPN, and TCP-flag
- Packet fields: source IP address, destination IP address, protocol, source port, destination port, source IP prefix, and destination IP prefix

Different services have different requirements for link quality indicators: delay, jitter, packet loss rate, and composite measure indicator (CMI). If services are insensitive to a link quality indicator, no threshold needs to be configured for this indicator.

Copyright © Huawei Technologies Co., Ltd.

NOTE

- On a device, the maximum delay is 5000 ms, the maximum jitter is 3000 ms, and the maximum packet loss rate is 1000‰. These maximum values are also the default values. Smaller values of the preceding indicators indicate better link quality.
- The CMI is calculated using the following formula: CMI = 9000 – *cmi-method*. The default value of *cmi-method* is the sum of delay (D), jitter (J), and packet loss rate (L).
  - The *cmi-method* parameter is configurable. For example, if *cmi-method* is set to D+10*J, the delay is 1000 ms, and the jitter is 10 ms, then the CMI is calculated as follows:

    CMI = 9000 – (D + 10 * J) = 9000 – (1000 + 100) = 7900
  - In the CMI calculation formula, if there is a coefficient before a link quality indicator (delay, jitter, or packet loss rate), the product of the link quality indicator and the coefficient must be less than or equal to the maximum value of the indicator. For example, if *cmi-method* is set to 10*D+10*J +10*L, the delay is 1000 ms, the jitter is 100 ms, and the packet loss rate is 10‰, then the CMI is calculated as follows:

    CMI = 9000 – (5000 + 1000 + 100) = 2900

    In the preceding formula, 10*D is 10000, which exceeds the maximum value of 5000. Therefore, the delay takes the maximum value 5000.

Detection Link and Link Group

SPR uses detection links to implement intelligent route selection. Each detection link has a unique probe, which is a network quality analysis (NQA) test instance. If an NQA test fails, the corresponding detection link is unavailable. A probe obtains quality indicators of the corresponding detection link. Then SPR selects an optimal link based on the link quality indicators.

SPR uses link groups instead of detection links to measure link quality. A detection link can be added to different link groups, and a link group can have one or more detection links.

SPR defines three roles for links: active link group, standby link group, and best-effort link. When no suitable link is available in the active and standby link groups, SPR activates the best-effort link to forward service data.

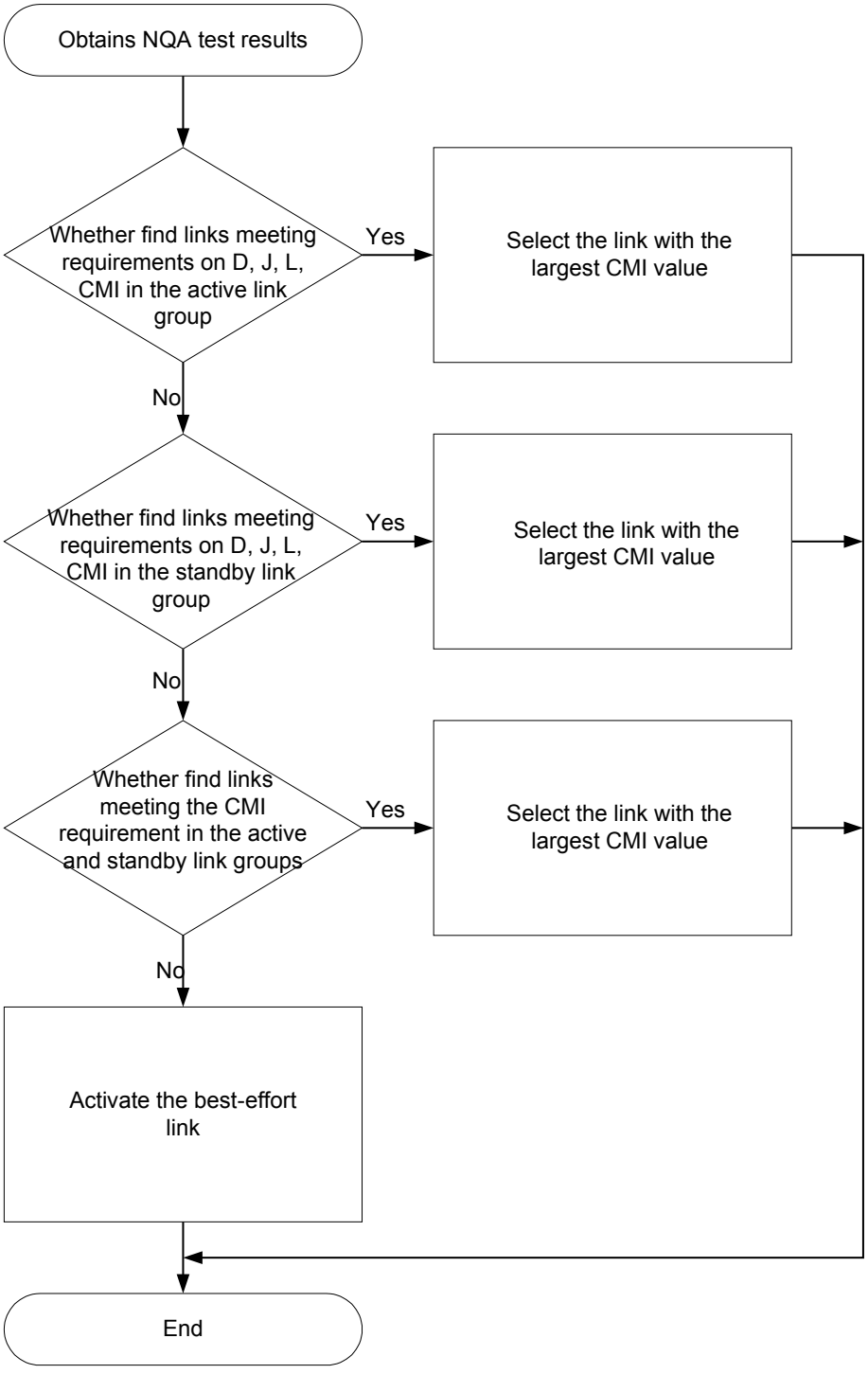
A link group can be bound to different services. For example, link group 1 can be used as the active link group for service 1 and as the standby link group for service 2.

Link Selection

Figure 8-1 shows how SPR selects links for services based on NQA test results.

Copyright © Huawei Technologies Co., Ltd.

Figure 8-1 Link selection by SPR



If a link is transmitting services, the probability that this link is selected to transmit new services is lower. This ensures load balancing among links.

Switchover Timer

SPR uses a switchover timer to control link switchovers when the link quality does not meet service requirements.

Copyright © Huawei Technologies Co., Ltd.

SPR periodically obtains NQA test results to determine whether a link meets service requirements. If SPR obtains NQA test results indicating that a link does not meet service requirements, SPR starts the switchover timer. Before the switchover timer expires:

- If SPR obtains NQA test results indicating that the link meets service requirements, SPR resets the timer.
- If SPR does not obtain NQA test results indicating that the link meets service requirements, SPR triggers a link switchover.

Flapping Suppression Timer

When a network is unstable, SPR triggers link switchovers frequently, which degrades service experience. SPR provides the flapping suppression function to address this problem.

The flapping suppression function is disabled by default, and the flapping suppression period is configurable. After traffic is switched to a new link, SPR starts the flapping suppression timer. Within a flapping suppression period, SPR does not perform a link switchover even if it does not obtain link NQA test results indicating that the link meets service requirements within a switchover period. After the flapping suppression timer expires:

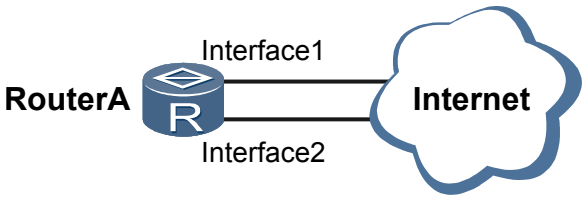
- If SPR still does not obtain NQA test results indicating that the link meets service requirements within a switchover period, SPR performs a link switchover.
- If SPR obtains NQA test results indicating that the link meets service requirements within a switchover period, SPR retains traffic on the link.

8.3 Applications

Local PBR

As shown in **Figure 8-2**, RouterA connects to the Internet using two links. To apply PBR to packets that are generated on RouterA, configure local PBR.

Figure 8-2 Networking diagram of local PBR

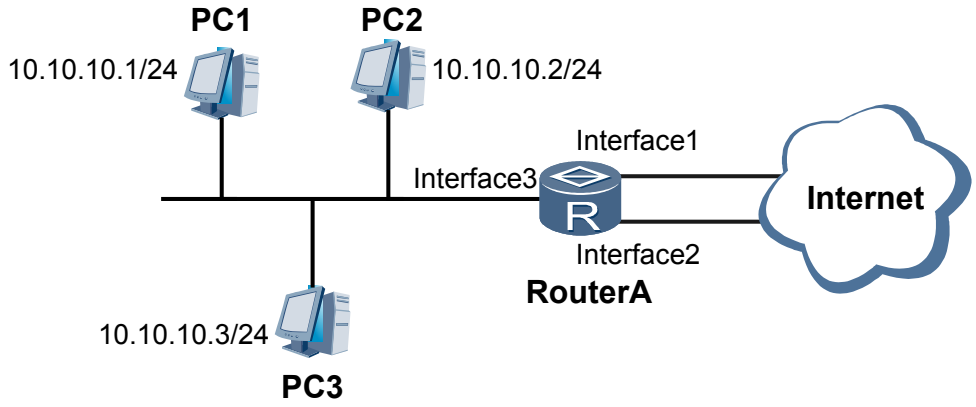


Interface PBR

As shown in **Figure 8-3**, the intranet connects to the Internet using RouterA. RouterA has two outbound interfaces connected to the Internet. To enable packets of a specified type to be forwarded through a specified outbound interface, configure interface PBR.

Copyright © Huawei Technologies Co., Ltd.

Figure 8-3 Networking diagram of interface PBR



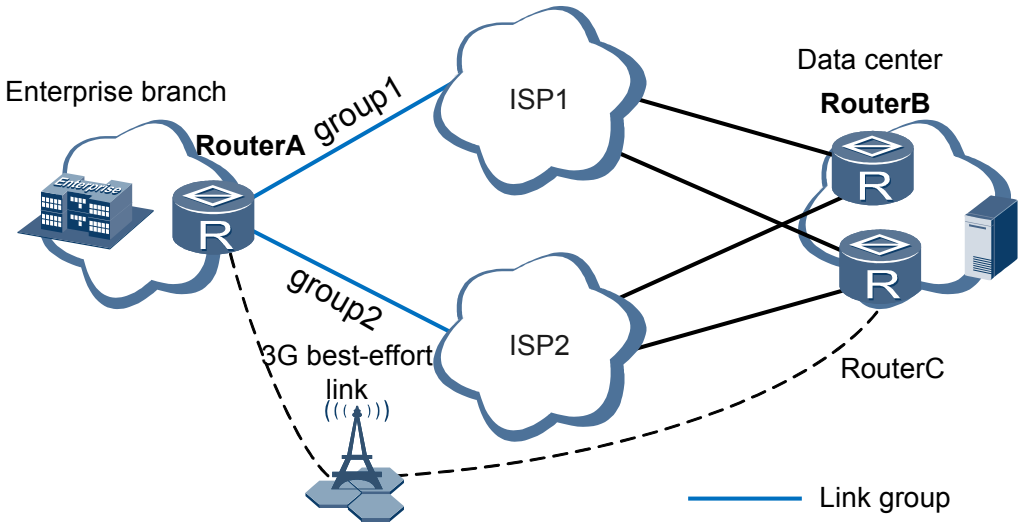
- Routing based on source addresses: Network administrators can use interface PBR to ensure that VIP users use the link with a higher rate and other users use the link with a lower rate. Interface PBR is configured on RouterA to define routing rules and actions. For example, interface PBR is enabled on Interface3. The interface PBR allows RouterA to send all packets that are received on Interface3 from PC1 at 10.10.10.1/24 through Interface2 and send other packets based on their destination addresses.
- Routing based on service classes: Different services have different requirements for the transmission rate, throughput, and reliability. Interface PBR allows routers to route data packets based on service classes and network status. For example, routers use high-bandwidth links to transmit voice and video services and low-bandwidth links to transmit data services. Assume that the bandwidth of the link for sending packets from Interface1 is higher than that from Interface2. Interface PBR can be configured on Interface3 of RouterA to enable RouterA to send voice and video services from Interface1 and data services from Interface2.

SPR

As shown in **Figure 8-4**, an enterprise branch connects to the enterprise data center over two ISP networks (ISP1 and ISP2), and a 3G outbound interface is configured on RouterA to provide a best-effort link. RouterA connects to ISP1 through the link group group1 and connects to ISP2 through the link group group2. ISP1 provides advanced network service at a high cost, while ISP2 provides common network service at a low cost. The enterprise branch exchanges voice, video, FTP and HTTP services with the data center. Voice and video services require high link quality. Therefore, group1 and group2 function as the active and standby link groups respectively for voice and video services. FTP and HTTP services do not require high link quality. Therefore, group2 and group1 function as the active and standby link groups respectively for FTP and HTTP services.

Copyright © Huawei Technologies Co., Ltd.

Figure 8-4 Networking diagram of SPR



SPR processes voice and video services as follows:

- When link quality provided by group1 cannot meet voice and video service requirements but link quality provided by group2 does, SPR switches voice and video services to group2 after the flapping suppression timer expires.
  - If group1 meets service requirements again, SPR switches services back to group1 after the flapping suppression timer expires and a switchover period elapses.
  - If both group 1 and group2 cannot meet service requirements later but group1 provides better link quality than group2, SPR switches services back to group1 after the flapping suppression timer expires and a switchover period elapses.
  - If group2 cannot meet service requirements later but provides better link quality than group1, SPR continues to use group2 to transmit services.
- When the link quality provided by group1 and group 2 cannot meet voice and video service requirements but their links are available, SPR selects the link with the largest CMI value from group1 and group2 to transmit services.
- When all the links in group1 and group2 are unavailable, SPR uses the 3G best-effort link to transmit services.

8.4 References

None

Copyright © Huawei Technologies Co., Ltd.