IPv6 技术白皮书

目 录

1 概述

IPv6（Internet Protocol Version 6，互联网协议版本 6）是网络层协议的第二代标准协议，也被称为 IPng（IP Next Generation，下一代互联网协议）。IPv6 不仅解决了 IPv4 地址空间不足的问题，还在 IPv4 协议的基础上进行了一些改进，例如，通过扩展头提高 IPv6 协议的可扩展性、内置安全性解决网络安全问题。

IPv6 可以为互联网和物联网提供更加广泛的连接，实现万物互联，打造数字化基础设施，促进物联网、工业互联网、人工智能等新应用、新领域的创新。在 5G、物联网等新兴领域飞速发展的今天，IPv6 协议的魅力不断展现，IPv6 协议获得了更加广阔的发展空间。

本文在讲解IPv6技术的优势、基于IPv6的应用协议扩展后，将介绍IPv6协议的发展方向（即IPv6+），并提供几种常见的 IPv6 部署方案，以帮助用户理解和部署 IPv6 协议。

2 IPv6 技术优势

2.1 充足的地址空间

IPv6 地址的长度是 128 比特（16 字节），可以提供超过 3.4×10^{38} 个地址。在万物互联的需求下，IPv6 具有足够大的地址空间，可以为每一个具有联网需求的终端提供 IPv6 地址，而不用担心地址耗尽，极大地增强了互联网的服务能力。

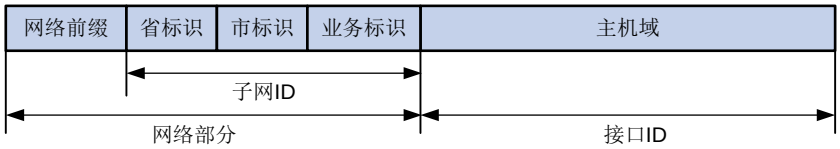
2.2 层次化的地址结构

IPv6 的地址空间采用了层次化的地址结构，地址管理更加便捷，且有利于路由快速查找，借助路由聚合，还可以有效减少 IPv6 路由表占用的系统资源。

IPv6 地址使用多层等级结构。地址注册机构分配 IPv6 地址范围后，服务提供商、组织机构可以根据各自的需要在该 IPv6 地址范围内分层级、更加精细地划分地址范围，以管理所辖范围内的地址分布。如图 1 所示，IPv6 地址由以下几部分组成：

- 网络前缀：由 CNNIC（China Internet Network Information Center，中国互联网络信息中心）和 ISP 分配。
- 子网 ID：组织机构根据需要分层级划分地址范围。例如，先根据地域分别为省、市分配地址范围，再按照业务类型分配地址范围。
- 接口 ID：网络中主机的标识。

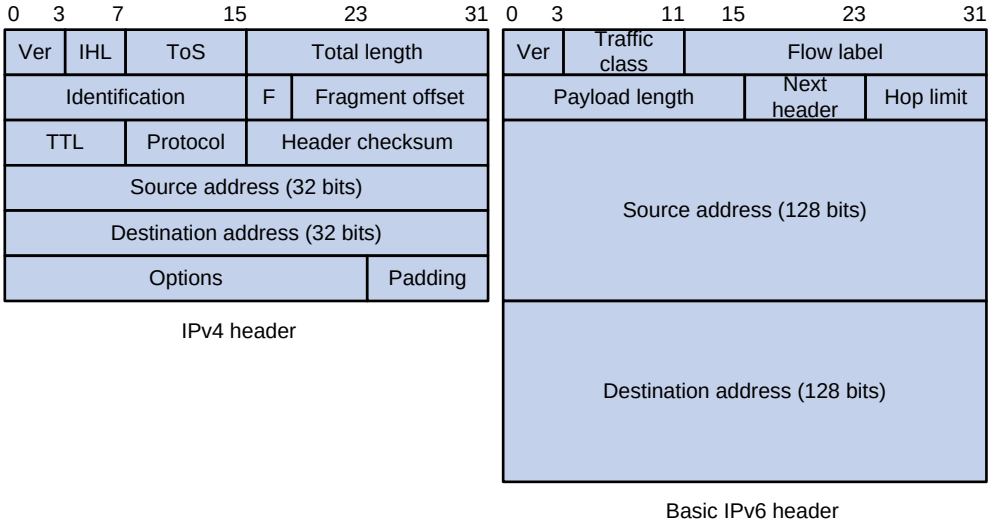
图1 IPv6 地址结构



2.3 简化报文头

通过将 IPv4 报文头中的某些字段裁减或移入到扩展报文头，减小了 IPv6 基本报文头的长度。IPv6 使用固定长度的基本报文头，从而简化了转发设备对 IPv6 报文的处理，提高了转发效率。尽管 IPv6 地址长度是 IPv4 地址长度的四倍，但 IPv6 基本报文头的长度只有 40 字节，为 IPv4 报文头长度（不包括选项字段）的两倍。

图2 IPv4 报文头和 IPv6 基本报文头格式比较



2.4 灵活的扩展头

IPv6 取消了 IPv4 报文头中的选项字段，并引入了多种扩展报文头，在提高处理效率的同时还大大增强了 IPv6 的灵活性，为 IP 协议提供了良好的扩展能力。IPv4 报文头中的选项字段最多只有 40 字节，而 IPv6 扩展报文头的大小只受到 IPv6 报文大小的限制。

2.4.1 IPv6 扩展头类型

IPv6 支持的扩展头如表 1 所示。扩展头使得 IPv6 协议具有良好的扩展性。根据业务需要，IPv6 不仅可以定义新的扩展头，还可以在已有扩展头中定义新的子扩展头。

表1 IPv6 扩展头

扩展头名称	类型值	处理节点	用途
逐跳选项头（Hop-by-Hop Options Header）	0	报文转发路径上的所有节点	用于巨型载荷告警、路由器告警、预留资源（RSVP）
路由头（Routing Header）	43	目的节点及报文必须经过的中间节点	用来指定报文必须经过的中间节点
分段头（Fragment Header）	44	目的节点	当IPv6报文的长度超过报文经过路径的PMTU（Path MTU，路径MTU）时，源节点将通过分段头对该IPv6报文进行分片 在IPv6中，仅源节点可以对报文进行分片，中

扩展头名称	类型值	处理节点	用途
			间节点不可以对报文进行分片 说明 PMTU 是从源节点到目的节点的报文转发路径上最小的 MTU
封装安全载荷头（Encapsulating Security Payload Header，ESP Header）	50	目的节点	用来提供数据加密、数据来源认证、数据完整性校验和抗重放功能
认证头（Authentication Header）	51	目的节点	用来提供数据来源认证、数据完整性校验和抗重放功能，它能保护报文免受篡改，但不能防止报文被窃听，适合用于传输非机密数据 AH提供的认证服务要强于ESP
目的选项头（Destination Options Header）	60	目的节点、路由头中指定的中间节点	用来携带传递给目的节点、路由头中指定中间节点的信息。例如，移动IPv6中，目的选项头可以用于在移动节点和家乡代理之间交互注册信息

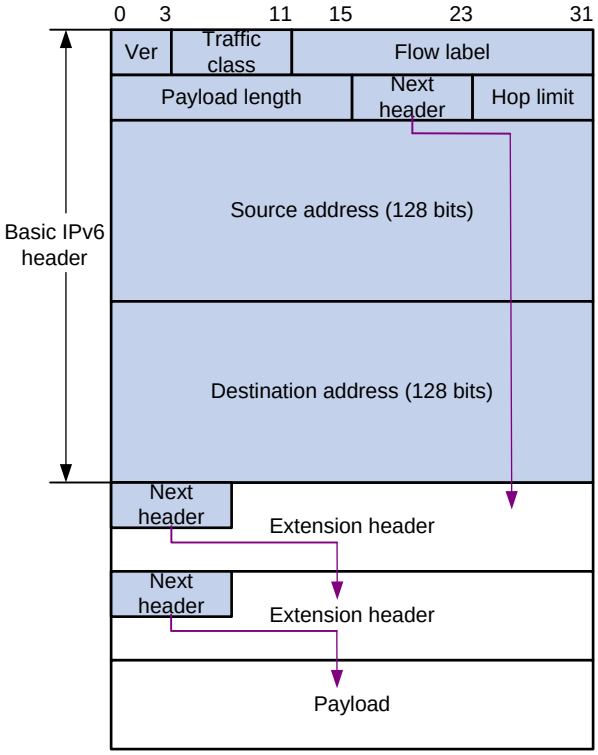
2.4.2 IPv6 扩展头的报文格式

一个 IPv6 报文可以携带 0 个、1 个或多个扩展头。如图 3 所示，IPv6 通过 Next header 字段标明下一个扩展头的类型。例如，IPv6 基本报文头中的 Next header 字段取值为 43 时，表示紧跟在 IPv6 基本报文头后的扩展头为路由头；路由头中的 Next header 字段取值为 44 时，表示路由头后的扩展头为分段头。

说明

最后一个扩展头的 Next header 字段用来标识 Payload 类型。例如，取值为 6，表示 Payload 为 TCP 报文；取值为 17，表示 Payload 为 UDP 报文。

图3 IPv6 扩展头的报文格式



2.5 强大的邻居发现协议

IPv6 的邻居发现协议是通过一组 ICMPv6（Internet Control Message Protocol for IPv6，IPv6 互联网控制消息协议）消息实现的，管理着邻居节点间（即同一链路上的节点）信息的交互。它代替了 ARP（Address Resolution Protocol，地址解析协议）、ICMPv4 路由器发现和 ICMPv4 重定向消息，并提供了一系列其他功能：

- 地址解析：获取同一链路上邻居节点的链路层地址，功能与 IPv4 的 ARP 相同。
- 邻居可达性检测：在获取到邻居节点的链路层地址后，检测邻居节点的状态，验证邻居节点是否可达。
- 重复地址检测：当节点获取到一个 IPv6 地址后，验证该地址是否已被其他节点使用，与 IPv4 的免费 ARP 功能相似。
- 路由器发现/前缀发现：节点获取邻居路由器的信息、所在网络的前缀、以及其他配置参数。节点获取到所在网络前缀后，可以根据该前缀自动生成 IPv6 地址，该过程称为 IPv6 地址无状态自动配置。
- 重定向功能：当网络中存在更优的路径时，路由器向主机发送 ICMPv6 重定向报文，通知主机选择更好的下一跳进行后续报文的发送。该功能与 IPv4 的 ICMP 重定向消息的功能相同。

2.6 内置安全性

IPv4 协议本身未提供加密、认证等安全功能，需要与其他安全协议（如 IPsec）配合使用，或由应用协议来提供安全性，增加了应用协议设计的复杂度。IPv6 协议在设计时便充分考虑了安全问题，

在 IPv6 协议中定义了 ESP 头和认证头，通过 ESP 头和认证头为报文传输提供端到端的安全性。基于 IPv6 协议的应用可以直接继承 IPv6 协议的安全功能，为解决网络安全问题提供了标准，并提高了不同 IPv6 应用之间的互操作性。

3 基于 IPv6 的协议扩展

用于设备管理的协议（Telnet、SSH、SNMP 等）、高可靠性机制（VRRP、M-LAG 等）、安全协议（802.1X、端口安全等）无需改动或稍做修改即可支持 IPv6。但是，还有一些 IPv4 网络中运行的应用层协议、路由协议、组播协议、安全协议等，为了适应 IPv6 协议，需要进行一些扩展。本节介绍这些协议在 IPv6 网络中的扩展方式。

3.1 IPv6全球单播地址配置方式

节点可以通过如下方式获取 IPv6 全球单播地址：

- 手工配置：用户手工为节点上的接口指定 IPv6 单播地址。
- 无状态地址自动配置：节点通过邻居发现协议获取到网络前缀后，根据该前缀自动生成 IPv6 单播地址。
- 有状态地址自动配置：节点通过 DHCPv6 协议从 DHCPv6 服务器获取 IPv6 单播地址。

不同 IPv6 全球单播地址配置方式的适用场景如表 2 所示。

表2 不同 IPv6 全球单播地址配置方式的适用场景

地址配置方式	优缺点	适用场景	前缀长度要求
手工配置	优点：无需协议报文交互 缺点：手工配置工作量，且无法动态调整	链路本地地址或Loopback接口地址	无要求，可自定义
无状态地址自动配置	优点：无需额外部署服务器，实现较为简单 缺点：无法精确控制为节点分配的IPv6地址	对终端访问行为无强管控需求。例如物联网终端（视频监控、路灯等）	固定为64位
有状态地址自动配置	优点：可以精确控制分配给节点的IPv6地址，并记录地址分配信息 缺点：需要在网络中部署DHCPv6服务器，实现较为复杂	对终端访问行为有强管控需求。例如校园网、办公区等	无要求，可自定义

无状态地址自动配置和有状态地址自动配置可以配合使用。例如，通过无状态地址自动配置获取 IPv6 地址后，使用有状态地址自动配置获取其他网络配置参数（如 DNS 服务器地址等）。

3.1.2 无状态地址自动配置

1. 无状态地址自动配置工作机制

无状态地址自动配置通过 IPv6 的邻居发现协议实现，其工作过程为：

(1) 路由器通过以下两种方式通告前缀信息：

- 路由器周期性地向所有节点的多播地址（FF02::1）发送 RA（Router Advertisement）消息，其中包括 IPv6 前缀、前缀的生命期、跳数限制等信息。
- 节点启动时，向所有路由器的多播地址（FF02::2）发送 RS（Router Solicitation）消息，路由器接收到 RS 消息后，向所有节点的多播地址（FF02::1）应答 RA 消息。

说明

前缀的生命期包括如下两种：

- 有效生命期：表示前缀有效期。在有效生命期内，通过该前缀自动生成的地址可以正常使用；有效生命期过期后，通过该前缀自动生成的地址变为无效，将被删除。
- 首选生命期：表示首选通过该前缀无状态自动配置地址的时间。首选生命期过期后，节点通过该前缀自动配置的地址将被废止。节点不能使用被废止的地址建立新的连接，但是仍可以接收目的地址为被废止地址的报文。首选生命期必须小于或等于有效生命期。

(2) 节点将路由器返回的 RA 消息中的地址前缀与本地的接口 ID 组合，生成 IPv6 单播地址。节点还会根据 RA 消息返回的配置信息自动配置节点，例如将 RA 消息中的跳数限制设置为本地发送的 IPv6 报文的最大跳数。

(3) 节点对生成的 IPv6 单播地址进行重复地址检测。检测方法为：节点在本地链路上，发送 NS（Neighbor Solicitation，邻居请求）消息，NS 消息的目的地址为根据前缀自动生成的 IPv6 单播地址的被请求节点多播地址。如果节点没有接收到 NA（Neighbor Advertisement，邻居通告）消息，则认为该地址不存在冲突，可以使用；否则，认为该地址存在冲突，不会使用该地址。

说明

被请求节点（Solicited-Node）多播地址主要用于获取同一链路上邻居节点的链路层地址及实现重复地址检测。每一个单播或任播 IPv6 地址都有一个对应的被请求节点地址。其格式为：

FF02:0:0:0:0:1:FFXX:XXXX

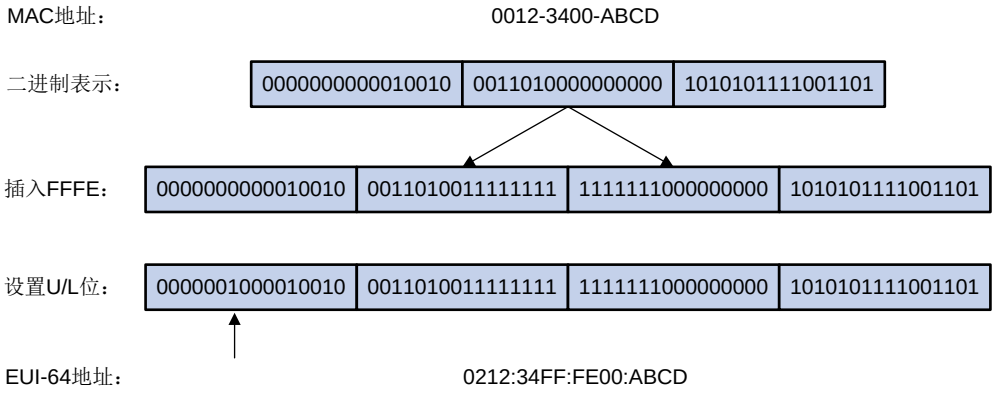
其中，FF02:0:0:0:0:1:FF 为 104 位固定格式；XX:XXXX 为单播或任播 IPv6 地址的后 24 位。

2. IEEE EUI-64 格式的接口 ID

节点自动根据本地信息生成借口 ID，不同接口的 IEEE EUI-64 格式接口 ID 的生成方法不同：

- 所有 IEEE 802 接口类型（例如，以太网接口、VLAN 接口）：IEEE EUI-64 格式的接口 ID 是从接口的链路层地址（MAC 地址）变化而来的。IPv6 地址中的接口 ID 是 64 位，而 MAC 地址是 48 位，因此需要在 MAC 地址的中间位置（从高位开始的第 24 位后）插入十六进制数 FFFE（1111111111111110）。为了使接口 ID 的作用范围与原 MAC 地址一致，还要将 Universal/Local (U/L)位（从高位开始的第 7 位）进行取反操作。最后得到的这组数就作为 EUI-64 格式的接口 ID。

图4 MAC 地址到 EUI-64 格式接口 ID 的转换过程



- Tunnel 接口：IEEE EUI-64 格式的接口 ID 的低 32 位为 Tunnel 接口的源 IPv4 地址，ISATAP 隧道的接口 ID 的高 32 位为 0000:5EFE，其他隧道的接口 ID 的高 32 位为全 0。
- 其他接口类型（例如，Serial 接口）：IEEE EUI-64 格式的接口 ID 由设备随机生成。

3. 自动生成接口 ID 不断变化的 IPv6 地址

接口根据无状态地址自动配置自动生成 IPv6 全球单播地址时，如果接口是 IEEE 802 类型的接口（例如，以太网接口、VLAN 接口），其接口 ID 是由 MAC 地址根据一定的规则生成，此接口 ID 具有全球唯一性。对于不同的前缀，接口 ID 部分始终不变，攻击者通过接口 ID 可以很方便地识别出通信流量是由哪台设备产生的，并分析其规律，会造成一定的安全隐患。

在无状态地址自动配置时，如果自动生成接口 ID 不断变化的 IPv6 地址，则可以加大攻击的难度，从而保护网络。为此，设备提供了临时地址功能，进行无状态地址自动配置，IEEE 802 类型的接口可以同时生成两类地址：

- 公共地址：地址前缀采用 RA 报文携带的前缀，接口 ID 由 MAC 地址产生。接口 ID 始终不变。
- 临时地址：地址前缀采用 RA 报文携带的前缀，接口 ID 由系统根据 MD5 算法计算产生。接口 ID 不断变化。

指定优先选择临时地址后，系统将优先选择临时地址作为报文的源地址。当临时地址的有效生命期过期后，这个临时地址将被删除，同时，系统会通过 MD5 算法重新生成一个接口 ID 不同的临时地址。所以，该接口发送报文的源地址的接口 ID 总是在不停变化。如果生成的临时地址因为 DAD 冲突不可用，就采用公共地址作为报文的源地址。

4. 前缀重新编址

当用户需要切换到新的网络前缀时，利用无状态地址自动配置可以方便、透明地实现前缀重新编址。前缀重新编址的过程为：

(1) 路由器在本地链路上通过 RA 消息发布旧的 IPv6 前缀，将该前缀的有效生命期和首选生命期降低到接近于 0 的值。

(2) 路由器在本地链路上通过 RA 消息发布新的 IPv6 前缀。

(3) 节点上同时存在新旧两个前缀生成的两个 IPv6 地址——新 IPv6 地址和旧 IPv6 地址。使用旧 IPv6 地址的连接仍然可以被处理，新建立的连接使用新 IPv6 地址。当旧 IPv6 地址的有效生命期结束后，该地址不再使用，节点仅使用新 IPv6 地址通信。

3.1.3 有状态地址自动配置（DHCPv6）

DHCPv6（Dynamic Host Configuration Protocol for IPv6，支持 IPv6 的动态主机配置协议）针对 IPv6 编址方案设计，用来为主机分配 IPv6 前缀、IPv6 地址和其他网络配置参数。

1. DHCPv6 的优点

与其他 IPv6 地址分配方式（手工配置、无状态地址自动配置）相比，DHCPv6 具有以下优点：

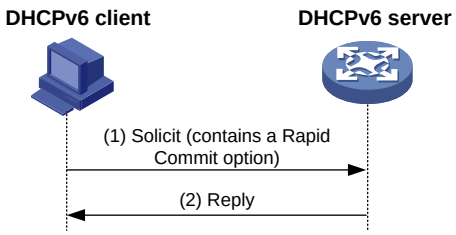
- 更好地控制地址的分配。通过 DHCPv6 不仅可以记录为主机分配的地址，还可以为特定主机分配特定的地址，以便于网络管理。
- 为 DHCPv6 客户端分配前缀，DHCPv6 客户端再作为路由器将前缀通告给主机，以便于主机无状态自动配置 IPv6 地址。通过这种方式，可以减少 DHCPv6 服务器管理的 IPv6 地址数量，并实现全网络的自动配置和管理。
- 除了 IPv6 前缀、IPv6 地址外，还可以为主机分配 DNS 服务器、域名后缀等网络配置参数。

2. DHCPv6 地址/前缀分配过程

DHCPv6 服务器为客户端分配地址/前缀的过程分为两类：

- 交互两个消息的快速分配过程

  图5 地址/前缀快速分配过程

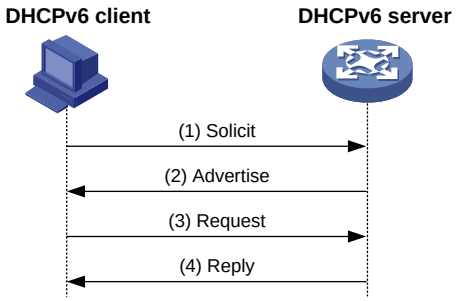
  如图 5 所示，地址/前缀快速分配过程为：

  a. DHCPv6 客户端在向 DHCPv6 服务器发送的 Solicit 消息中携带 Rapid Commit 选项，标识客户端希望服务器能够快速为其分配地址/前缀和其他网络配置参数。

  b. 如果 DHCPv6 服务器支持快速分配过程，则直接返回 Reply 消息，为客户端分配 IPv6 地址/前缀和其他网络配置参数。如果 DHCPv6 服务器不支持快速分配过程，则采用交互四个消息的分配过程为客户端分配 IPv6 地址/前缀和其他网络配置参数。

- 交互四个消息的分配过程

  图6 交互四个消息的分配过程

交互四个消息分配过程的简述如表 3。

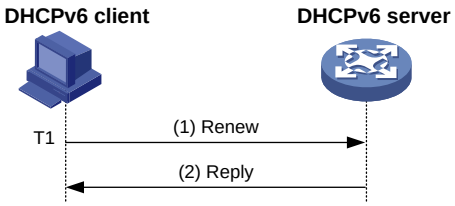
表3 交互四个消息的分配过程

步骤	发送的消息	说明
(1)	Solicit	DHCPv6客户端发送该消息，请求DHCPv6服务器为其分配IPv6地址/前缀和网络配置参数
(2)	Advertise	如果Solicit消息中没有携带Rapid Commit选项，或Solicit消息中携带Rapid Commit选项，但服务器不支持快速分配过程，则DHCPv6服务器回复该消息，通知客户端可以为其分配的地址/前缀和网络配置参数
(3)	Request	如果DHCPv6客户端接收到多个服务器回复的Advertise消息，则根据消息接收的先后顺序、服务器优先级等，选择其中一台服务器，并向该服务器发送Request消息，请求服务器确认为其分配地址/前缀和网络配置参数
(4)	Reply	DHCPv6服务器回复该消息，确认将地址/前缀和网络配置参数分配给客户端使用

3. 地址/前缀租约更新过程

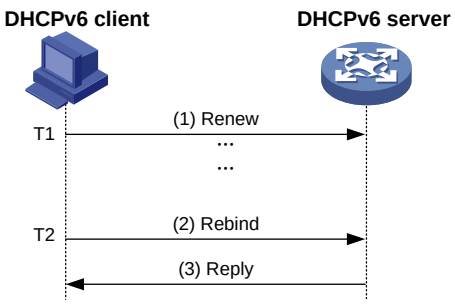
DHCPv6 服务器分配给客户端的 IPv6 地址/前缀具有一定的租借期限，该租借期限称为租约。租借期限由有效生命期决定。地址/前缀的租借时间到达有效生命期后，DHCPv6 客户端不能再使用该地址/前缀。在有效生命期到达之前，如果 DHCPv6 客户端希望继续使用该地址/前缀，则需要申请延长地址/前缀租约。

图7 通过 Renew 更新地址/前缀租约



如图 7 所示，地址/前缀租借时间到达时间 T1（推荐值为首选生命期的一半）时，DHCPv6 客户端会向为它分配地址/前缀的 DHCPv6 服务器发送 Renew 报文，以进行地址/前缀租约的更新。如果客户端可以继续使用该地址/前缀，则 DHCPv6 服务器回应续约成功的 Reply 报文，通知 DHCPv6 客户端已经成功更新地址/前缀租约；如果该地址/前缀不可以再分配给该客户端，则 DHCPv6 服务器回应续约失败的 Reply 报文，通知客户端不能获得新的租约。

图8 通过 Rebind 更新地址/前缀租约



如图 8 所示，如果在 T1 时发送 Renew 请求更新租约，但是未收到 DHCPv6 服务器的回应报文，则 DHCPv6 客户端会在 T2（推荐值为首选生命期的 0.8 倍）时，向所有 DHCPv6 服务器组播发送 Rebind 报文请求更新租约。如果客户端可以继续使用该地址/前缀，则 DHCPv6 服务器回应续约成功的 Reply 报文，通知 DHCPv6 客户端已经成功更新地址/前缀租约；如果该地址/前缀不可以再分配给该客户端，则 DHCPv6 服务器回应续约失败的 Reply 报文，通知客户端不能获得新的租约；如果 DHCPv6 客户端未收到服务器的应答报文，则到达有效生命期后，客户端停止使用该地址/前缀。

4. DHCPv6 选项介绍

- Option 17

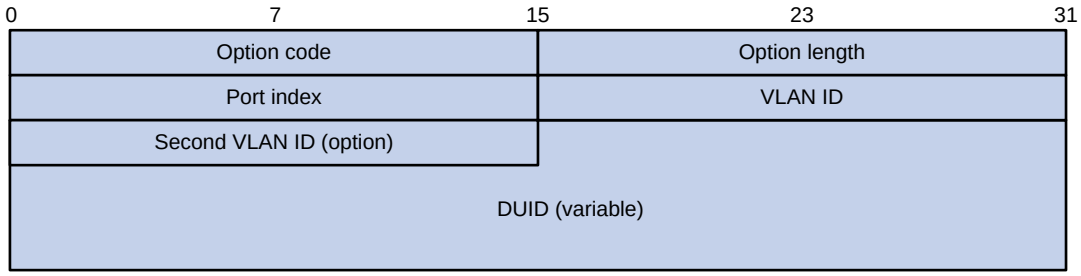
  Option 17 称为厂商自定义选项（Vendor-specific Information），是 RFC 中规定的保留选项，设备作为 DHCPv6 服务器时，可以利用该选项携带额外的网络参数（例如 TFTP 服务器名称、地址或设备的配置文件名等）并发送给 DHCPv6 客户端，以便为 DHCPv6 客户端提供相应的服务。

  为了提供可扩展性，通过 Option 17 为客户端分配更多的信息，Option 17 采用子选项的形式，通过不同的子选项为用户分配不同的网络配置参数，目前每个厂商自定义选项下最多配置 16 个子选项内容。

- Option 18

  Option 18 称为接口 ID 选项（Interface ID），设备接收到 DHCPv6 客户端发送的 DHCPv6 请求报文后，在该报文中添加 Option 18 选项（DHCPv6 中继会在 Relay-forwad 报文中添加 Option 18 选项），并转发给 DHCPv6 服务器。服务器可根据 Option 18 选项中的客户端信息选择合适的地址池为 DHCPv6 客户端分配 IPv6 地址。图 9 为 Option 18 选项格式。

  图9 Option 18 选项格式

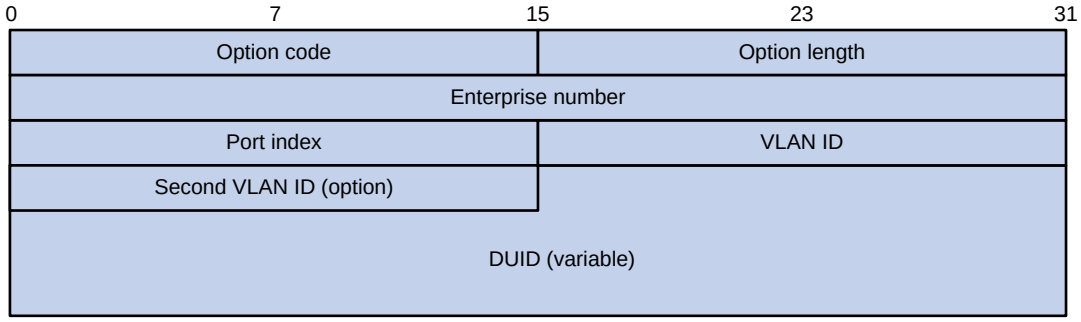
  各字段的解释如下：

  - Option code：Option 编号，取值为 18。
  - Option length：Option 字段长度。
  - Port index：DHCPv6 设备收到客户端请求报文的端口索引。
  - VLAN ID：第一层 VLAN 信息。
  - Second VLAN ID：第二层 VLAN 信息。选项格式中的 Second VLAN ID 字段为可选，如果 DHCPv6 报文中不含有 Second VLAN，则 Option 18 中也不包含 Second VLAN ID 内容。
  - DUID：缺省为设备本身的 DUID 信息，可通过命令行配置为其它 DUID 信息。

- Option 37

Option 37 称为远程 ID 选项（Remote ID），设备接收到 DHCPv6 客户端发送的 DHCPv6 请求报文后，在该报文中添加 Option 37 选项（DHCPv6 中继会在 Relay-forwad 报文中添加 Option 37 选项），并转发给 DHCPv6 服务器。服务器可根据 Option 37 选项中的信息对 DHCPv6 客户端定位，为分配 IPv6 地址提供帮助。图 10 为 Option 37 选项格式。

图10 Option 37 选项格式



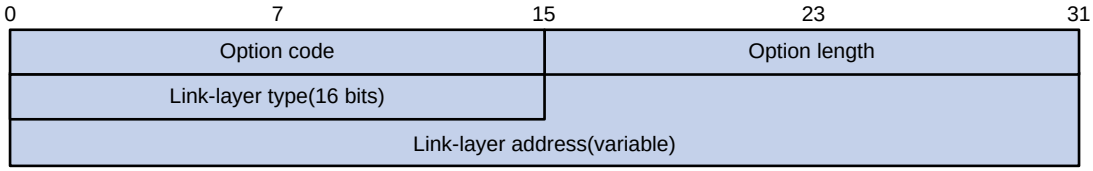
各字段的解释如下：

- Option code：Option 编号，取值为 37。
- Option length：Option 字段长度。
- Enterprise number：企业编号。
- Port index：DHCPv6 设备收到客户端请求报文的端口索引。
- VLAN ID：第一层 VLAN 信息。
- Second VLAN ID：第二层 VLAN 信息。选项格式中的 Second VLAN ID 字段为可选，如果 DHCPv6 报文中不含有 Second VLAN，则 Option 37 中也不包含 Second VLAN ID 内容。
- DUID：缺省为设备本身的 DUID 信息，可通过命令行配置为其它 DUID 信息。

- Option 79

Option 79 称为客户端链路地址选项（Client link layer address）。DHCPv6 请求报文经过第一个 DHCPv6 中继时，该 DHCPv6 中继会学习报文的源 MAC 地址，即 DHCPv6 客户端的 MAC 地址。DHCPv6 中继生成和请求报文对应的 Relay-Forward 报文时，会将学到的 MAC 地址添加到报文的 Option 79 选项中，再将该报文转发给 DHCPv6 服务器。DHCPv6 服务器可根据 Option 79 选项中的信息学习 DHCPv6 客户端的 MAC 地址，为 IPv6 地址/IPv6 前缀分配或客户端合法性认证提供帮助。图 11 为 Option 79 选项格式。

图11 Option 79 选项格式



各字段的解释如下：

- Option code：Option 编号，取值为 79。

- Option length：Option 字段长度。
- Link-layer type：客户端链路层地址类型。
- Link-layer address：客户端链路层地址。

3.2 IPv6 DNS

3.2.1 IPv6 DNS 简介

DNS（Domain Name System，域名系统）是一种用于 TCP/IP 应用程序的分布式数据库，提供域名与 IP 地址之间的转换。通过域名系统，用户进行某些应用时，可以直接使用便于记忆的、有意义的域名，而由网络中的域名解析服务器将域名解析为正确的 IP 地址。

在 IPv6 网络中，DNS 主要使用 AAAA 和 PTR 记录来实现域名与 IPv6 地址的转换。

- AAAA 记录：用来将域名映射为 IPv6 地址，实现正向地址解析。
- PTR 记录：用来将 IPv6 地址映射为域名，实现反向地址解析。

3.2.2 天窗问题避免

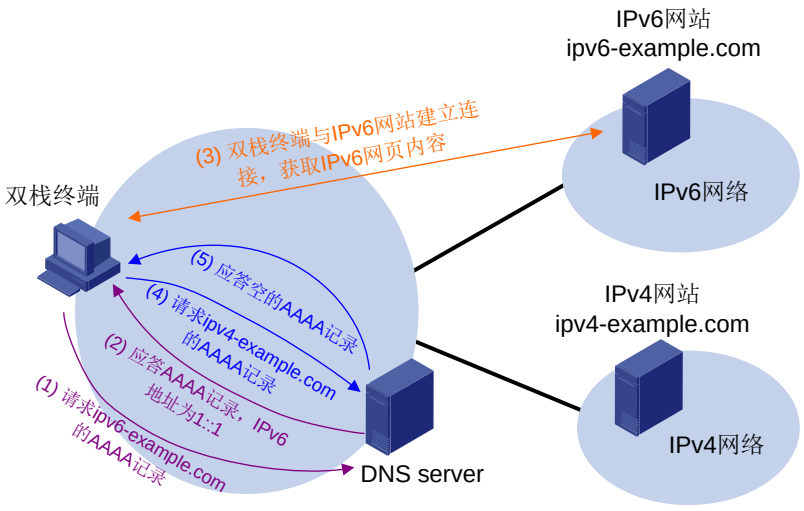
在 IPv4 网络向 IPv6 网络迁移的过程中，IPv4 网络和 IPv6 网络在一段时期内将共存。IPv4 和 IPv6 共存网络中，用户访问 IPv6 网页时，可能会出现天窗问题。

天窗问题是指用户访问 IPv6 网页时，如果网页中包含其他网站的链接（简称为外链），且该网站属于未进行 IPv6 改造升级的 IPv4 网络，则 IPv6 网页访问会出现响应缓慢、部分内容无法显示、部分功能无法使用等情况。

1. 避免双栈终端访问 IPv6 网页的天窗问题

如图 12 所示，天窗问题出现的原因为：双栈终端访问 IPv6 网页时，对于其中包含的 IPv4 网站链接，双栈终端将其视为 IPv6 网址，向 DNS 服务器请求 AAAA 记录。由于 IPv4 网站不存在对应的 AAAA 记录，导致域名解析失败。

图12 天窗问题产生的原因示意图



可以通过 IPv6 域名解析失败、IPv6 连接建立失败后，尝试进行 IPv4 域名解析、建立 IPv4 连接的方式解决天窗问题。具体工作过程为：

(1) 双栈终端优先发送 IPv6 DNS 请求，请求 IPv6 网页内嵌域名的 AAAA 记录。

(2) 双栈终端随后发送 IPv4 DNS 请求，请求 IPv6 网页内嵌域名的 A 记录。

(3) 如果双栈终端接收到域名服务器回复的 AAAA 记录，则双栈终端向 AAAA 记录中的 IPv6 地址发送连接建立请求，与内嵌网站建立 IPv6 连接。

(4) 如果双栈终端未收到 AAAA 记录，或与 AAAA 记录中的 IPv6 地址建立连接失败，则双栈终端向 A 记录中的 IPv4 地址发送连接建立请求，与内嵌网站建立 IPv4 连接。

2. 避免 IPv6 单栈终端访问 IPv6 网页的天窗问题

IPv6 单栈终端访问 IPv6 网页时，对于其中包含的 IPv4 网站链接，客户端浏览器会向本地 DNS 服务器发送请求外链域名的 IPv6 DNS 请求报文。由于 DNS 服务器不存在 IPv4 网站链接的 AAAA 记录，域名解析失败，进而出现天窗问题。为了解决上述场景中的天窗问题，需要部署 LB 设备并在 LB 设备上配置外链代理功能。

如图 13 所示，在 LB 设备上配置外链代理功能后，LB 设备在响应 IPv6 客户端访问主站页面的请求时，会同时向客户端返回外链改写的脚本文件。客户端浏览器执行该脚本文件，对 IPv4 外链域名进行改写。然后，客户端向本地 DNS 服务器查询改写后的外链域名。本地 DNS 服务器根据查询的域名将 DNS 请求重定向至 LB 设备，由 LB 设备代替 IPv6 客户端请求外链资源，并将外链资源返回给客户端。具体工作过程为：

(1) IPv6 Host 浏览器执行脚本文件，将外链域名改写为 http://www.extlink.com.proxy.suffix.com。

(2) IPv6 Host 向 Local DNS server 发送查询域名 http://www.extlink.com.proxy.suffix.com 的 DNS 请求报文。Local DNS server 根据查询结果通知 IPv6 Host 解析该域名的权威 DNS 服务器为 LB device。

(3) IPv6 Host 向 LB device 发送查询域名 http://www.extlink.com.proxy.suffix.com 的 DNS 请求报文。

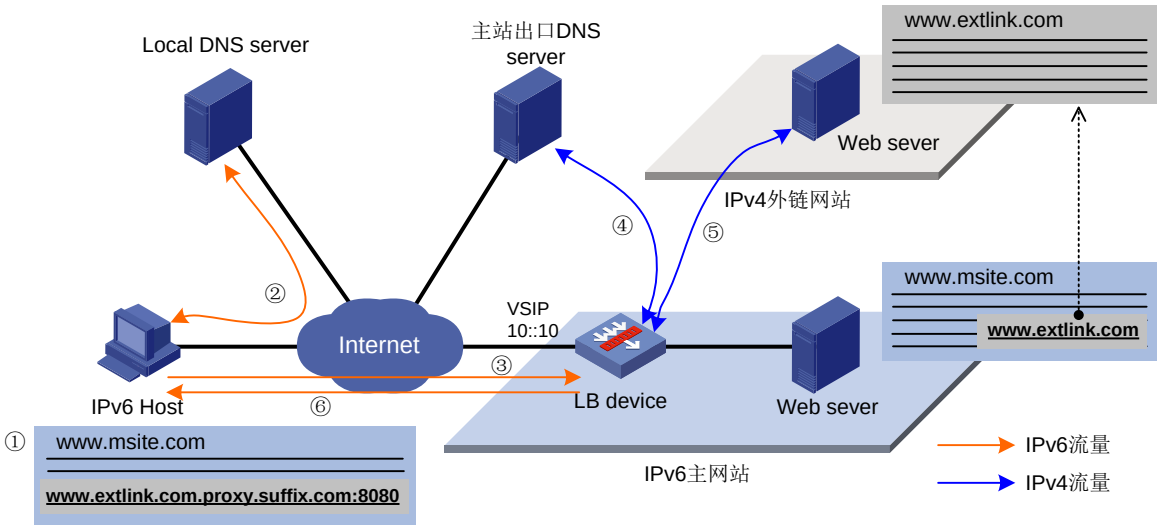
(4) LB device 收到包含改写后域名的 DNS 请求报文后，代替 IPv6 Host 向 DNS sever 发送查询原始外链域名 http://www.extlink.com 的 DNS 请求报文，获取外链资源的 IPv4 地址。

(5) LB device 根据外链资源的 IPv4 地址获取外链资源。

(6) LB device 将收到的外链资源发送给 IPv6 Host。

浏览器解析获取到的外链资源即可将正常的网页展示给用户。

图13 外链代理流程图



3.3 IPv6路由

IPv4 网络中常见的路由协议包括 RIP、OSPF、IS-IS 和 BGP。这些路由协议需要进行一定的演变和扩展才能应用于 IPv6 网络。扩展后的路由协议称为 RIPng、OSPFv3、IPv6 IS-IS 和 IPv6 BGP。IPv4 网络和 IPv6 网络的路由协议在应用场景、路由思路、优劣势等方面并无本质区别，只是为了适应 IPv6 地址及 IPv6 网络特点，调整了部分路由工作机制。

3.3.1 RIPng

RIP 有两个版本：RIP-1 和 RIP-2。

RIP-1 是有类别路由协议（Classful Routing Protocol），它只支持以广播方式发布协议报文。RIP-1 的协议报文无法携带掩码信息，它只能识别 A、B、C 类这样的自然网段的路由，因此 RIP-1 不支持不连续子网。

RIP-2 是一种无类别路由协议（Classless Routing Protocol），与 RIP-1 相比，它有以下优势：

- 支持路由标记，在路由策略中可根据路由标记对路由进行灵活的控制。
- 报文中携带掩码信息，支持路由聚合和 CIDR。
- 支持指定下一跳，在广播网上可以选择到最优下一跳地址。
- 支持组播方式发送更新报文，减少资源消耗。
- 在路由更新报文中增加一个认证 RTE（Route Entries，路由表项）以支持对协议报文进行验证，并提供明文验证和 MD5 验证两种方式，增强安全性。

RIPng 在工作机制上与 RIP-2 基本相同，但为了支持 IPv6 地址格式，RIPng 对 RIP-2 做了一些改动。

1. 报文的不同

- 路由信息中的目的地址和下一跳地址长度不同。

  RIP-2 报文中路由信息的目的地址和下一跳地址只有 32 比特，而 RIPng 均为 128 比特。

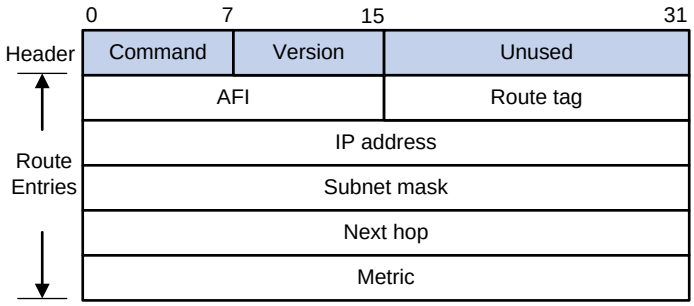
- 报文长度不同。

  RIP-2 对报文的长度有限制，规定每个报文最多只能携带 25 个 RTE，而 RIPng 对报文长度、RTE 的数目都不作规定，报文的长度与发送接口设置的 IPv6 MTU 有关。

- 报文格式不同。

  RIP-2 报文结构如图 14 所示，由头部（Header）和多个 RTE 组成。

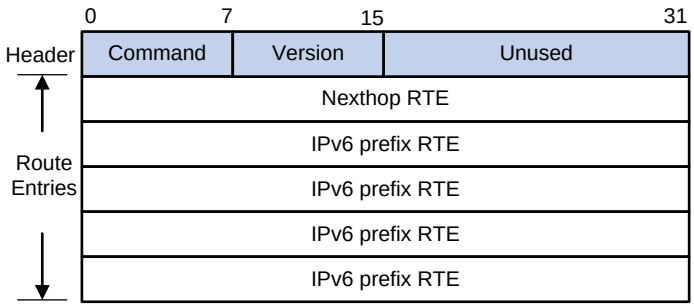
  图14 RIP-2 报文



RIPng 报文结构如图 15 所示。与 RIP-2 一样，RIPng 报文也是由 Header 和多个 RTE 组成。与 RIP-2 不同的是，在 RIPng 里有两类 RTE，分别是：

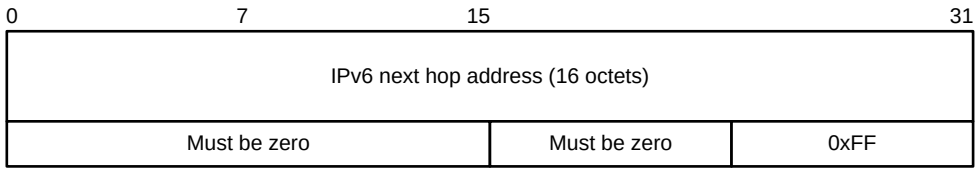
- 下一跳 RTE：位于一组具有相同下一跳的 IPv6 前缀 RTE 的前面，它定义了下一跳的 IPv6 地址。
- IPv6 前缀 RTE：位于某个下一跳 RTE 的后面。同一个下一跳 RTE 的后面可以有多个不同的 IPv6 前缀 RTE。它描述了 RIPng 路由表中的目的 IPv6 地址、路由标记、前缀长度以及度量值。

图15 RIPng 报文



下一跳 RTE 的格式如图 16 所示，其中，IPv6 next hop address 表示下一跳的 IPv6 地址。

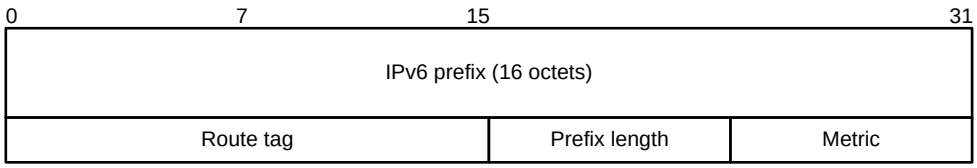
图16 下一跳 RTE 格式



IPv6 前缀 RTE 的格式如图 17 所示，各字段的解释如下：

- IPv6 prefix：目的 IPv6 地址的前缀。
- Route tag：路由标记。
- Prefix lenth：IPv6 地址的前缀长度。
- Metric：路由的度量值。

图17 IPv6 前缀 RTE 格式



- 报文的发送方式不同。

  RIP-2 可以根据用户配置采用广播或组播方式来周期性地发送路由信息；RIPng 使用组播方式周期性地发送路由信息。

2. 安全认证不同

RIPng 自身不提供认证功能，而是通过使用 IPv6 提供的安全机制来保证自身报文的合法性。因此，RIP-2 报文中的认证 RTE 在 RIPng 报文中被取消。

3. 与网络层协议的兼容性不同

RIP-2 不仅能在 IP 网络中运行，也能在 IPX 网络中运行；RIPng 只能在 IPv6 网络中运行。

3.3.2 OSPFv3

OSPFv3 在工作机制上与 OSPFv2 基本相同，但为了支持 IPv6 地址格式，OSPFv3 对 OSPFv2 做了一些改动。

1. OSPFv3 与 OSPFv2 的相同点

OSPFv3 在协议设计思路和工作机制与 OSPFv2 基本一致：

- 报文类型相同：包含 Hello、DD、LSR、LSU、LSAck 五种类型的报文。
- 区域划分相同。
- LSA 泛洪和同步机制相同：为了保证 LSDB 内容的正确性，需要保证 LSA 的可靠泛洪和同步。
- 路由计算方法相同：采用最短路径优先算法计算路由。
- 网络类型相同：支持广播、NBMA、P2MP 和 P2P 四种网络类型。
- 邻居发现和邻接关系形成机制相同：OSPF 路由器启动后，便会通过 OSPF 接口向外发送 Hello 报文，收到 Hello 报文的 OSPF 路由器会检查报文中所定义的参数，如果双方一致就会形成邻居关系。形成邻居关系的双方不一定都能形成邻接关系，这要根据网络类型而定，只有当双方成功交换 DD 报文，交换 LSA 并达到 LSDB 的同步之后，才形成真正意义上的邻接关系。
- DR 选举机制相同：在 NBMA 和广播网络中需要选举 DR 和 BDR。

2. OSPFv3 与 OSPFv2 的不同点

为了支持在 IPv6 环境中运行，指导 IPv6 报文的转发，OSPFv3 对 OSPFv2 做出了一些必要的改进，使得 OSPFv3 可以独立于网络层协议，而且只要稍加扩展，就可以适应各种协议，为未来可能的扩展预留了充分的可能。

OSPFv3 与 OSPFv2 不同主要表现在：

- 基于链路的运行。

  OSPFv2 是基于网络运行的，两个路由器要形成邻居关系必须在同一个网段。

  OSPFv3 的实现是基于链路，一条链路可以包含多个子网，节点即使不在同一个子网内，只要在同一链路上就可以直接通信。

- 使用链路本地地址。

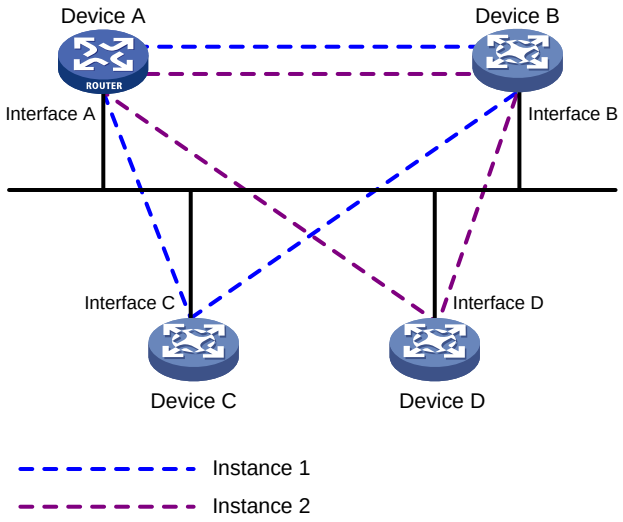
  OSPFv3 的路由器使用链路本地地址作为发送报文的源地址。一台路由器可以学习到这条链路上相连的所有其它路由器的链路本地地址，并使用这些链路本地地址作为下一跳来转发报文。但是在虚连接上，必须使用全球范围地址作为 OSPFv3 协议报文的源地址。

  由于链路本地地址只在本链路上有意义且只能在本链路上泛洪，因此链路本地地址只能出现在 Link LSA 中。

- 链路支持多实例复用。

  如图 18 所示，OSPFv3 支持在同一链路上运行多个实例，实现链路复用并节约成本。

  图18 链路支持多实例复用示意图



  Device A、Device B、Device C 和 Device D 连接到同一个广播网上，它们共享同一条链路。在 Device A 的 Interface A、Device B 的 Interface B、Device C 的 Interface C 上指定实例 1；在 Device A 的 Interface A、Device B 的 Interface B、Device D 的 Interface D 上指定实例 2，实现了 Device A、Device B 和 Device C 可以建立邻居关系，Device A、Device B 和 Device D 可以建立邻居关系。

  这是通过在 OSPFv3 报文头中添加 Instance ID 字段来实现的。如果接口配置的 Instance ID 与接收的 OSPFv3 报文的 Instance ID 不匹配，则丢弃该报文，从而无法建立邻居关系。

- 通过 Router ID 唯一标识邻居。

在 OSPFv2 中，当网络类型为点到点或者通过虚连接与邻居相连时，通过 Router ID 来标识邻居路由器，当网络类型为广播或 NBMA 时，通过邻居接口的 IP 地址来标识邻居路由器。

OSPFv3 取消了这种复杂性，无论对于何种网络类型，都是通过 Router ID 来唯一标识邻居。

- 认证的变化。

  OSPFv3 协议除了自身可以提供认证功能外，还可以通过使用 IPv6 提供的安全机制来保证自身报文的合法性。

- Stub 区域的支持。

  由于 OSPFv3 支持对未知类型 LSA 的泛洪，为防止大量未知类型 LSA 泛洪进入 Stub 区域，对于向 Stub 区域泛洪的未知类型 LSA 进行了明确规定：只有当未知类型 LSA 的泛洪范围是区域或链路而且 U 比特没有置位时，未知类型 LSA 才可以向 Stub 区域泛洪。

- 报文的不同。

  OSPFv3 报文封装在 IPv6 报文中，每一种类型的报文均以一个 16 字节的报文头部开始。

  与 OSPFv2 一样，OSPFv3 的五种报文都有同样的报文头，只是报文中的字段有些不同。

  OSPFv3 的 LSU 和 LSAck 报文与 OSPFv2 相比没有什么变化，但 OSPFv3 的报文头、Hello 与 OSPFv2 略有不同，报文的改变包括以下几点：

  - 版本号从 2 升级到 3。
  - 报文头的不同：与 OSPFv2 报文头相比，OSPFv3 报文头长度只有 16 字节，去掉了认证字段，但增加了 Instance ID 字段。Instance ID 字段用来支持在同一条链路上运行多个实例，且只在链路本地范围内有效。
  - Hello 报文的不同：与 OSPFv2 Hello 报文相比，OSPFv3 Hello 报文去掉了网络掩码字段，增加了 Interface ID 字段，用来标识发送该 Hello 报文的接口 ID。

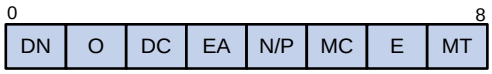
- Option 字段不同。

  在 OSPFv2 中，Option 字段出现在每一个 Hello 报文、DD 报文以及每一个 LSA 中。

  在 OSPFv3 中，Option 字段只在 Hello 报文、DD 报文、Router LSA、Network LSA、Inter Area Router LSA 以及 Link LSA 中出现。

  OSPFv2 的 Option 字段如图 19 所示。

  图19 OSPFv2 Option 字段格式



  OSPFv3 的 Option 字段如图 20 所示。

  图20 OSPFv3 Option 字段格式

  从上图可以看出，与 OSPFv2 相比，OSPFv3 的 Option 字段增加了 R 比特、V 比特。其中：

  - R 比特：用来标识设备是否是具备转发能力的路由器。如果 R 比特置 0，则表示该节点的路由信息将不会参加路由计算。如果当前设备不想转发目的地址不是本地地址的报文，可以将 R 比特置 0。

- V 比特：如果 V 比特置 0，该路由器或链路不会参加路由计算。

- LSA 类型格式不同。

  OSPFv3 支持七种类型的 LSA。OSPFv3 LSA 与 OSPFv2 LSA 的异同如表 4 所示。

  表4 OSPFv3 与 OSPFv2 LSA 的异同点

OSPFv2 LSA	OSPFv3 LSA	与 OSPFv2 LSA 异同点说明
Router LSA	Router LSA	名称相同，作用也类似，但是不再描述地址信息，仅仅用来描述路由域的拓扑结构
Network LSA	Network LSA	
Network Summary LSA	Inter Area Prefix LSA	作用类似，名称不同
ASBR Summary LSA	Inter Area Router LSA	
AS External LSA	AS External LSA	作用与名称完全相同
无	Link LSA	新增LSA
	Intra Area Prefix LSA	新增LSA

  OSPFv3 新增了 Link LSA 和 Intra Area Prefix LSA。

  - Router LSA 不再包含地址信息，使能 OSPFv3 的路由器为它所连接的每条链路产生单独的 Link LSA，将当前接口的链路本地地址以及路由器在这条链路上的一系列 IPv6 地址信息向该链路上的所有其它路由器通告。
  - Router LSA 和 Network LSA 中不再包含路由信息，这两类 LSA 中所携带的路由信息由 Intra Area Prefix LSA 来描述，该类 LSA 用来公告一个或多个 IPv6 地址前缀。

- LSA 处理方式不同。

  OSPFv3 扩大了 LSA 的泛洪范围。LSA 的泛洪范围已经被明确地定义在 LSA 的 LS Type 字段。目前，有三种 LSA 泛洪范围：

  - 链路本地范围：LSA 只在本地链路上泛洪，不会超出这个范围，该范围适用于新定义的 Link LSA。
  - 区域范围：LSA 的泛洪范围仅仅覆盖一个单独的 OSPFv3 区域。Router LSA、Network LSA、Inter Area Prefix LSA、Inter Area Router LSA 和 Intra Area Prefix LSA 都是区域范围泛洪的 LSA。
  - 自治系统范围：LSA 将被泛洪到整个路由域，AS External LSA 就是自治系统范围泛洪的 LSA。

  支持对未知类型 LSA 的处理方式不同。在 OSPFv2 中，收到类型未知的 LSA 将直接丢弃。OSPFv3 在 LSA 的 LS Type 字段中增加了一个 U 比特位来位标识对未知类型 LSA 的处理方式：

  - 如果 U 比特置 1，则对于未知类型的 LSA 按照 LSA 中的 LS Type 字段描述的泛洪范围进行泛洪。
  - 如果 U 比特置 0，对于未知类型的 LSA 仅在链路范围内泛洪。

- LSA 格式不同。

  为了适应 IPv6 地址长度和地址类型等需求，OSPFv3 对 LSA 头及各类 LSA 的格式进行了调整，详细介绍请参见《OSPFv3 技术白皮书》。

3.3.3 IPv6 IS-IS

为了支持在IPv6环境中运行，指导IPv6报文的转发，IPv6 IS-IS采用NLPID（Network Layer Protocol Identifier，网络层协议标识符）值142（0x8E）来标识IPv6协议，并通过对IS-IS TLV进行简单的扩展，使其能够处理IPv6的路由信息。

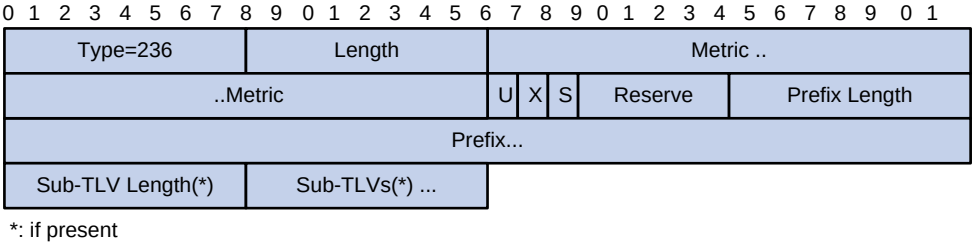
1. IPv6 IS-IS 新增 TLV

TLV（Type-Length-Value）是LSP（Link State PDU，链路状态协议数据单元）中的一个可变长字段值。为了支持IPv6路由的处理和计算，IS-IS新增了两个TLV，分别是：

- IPv6可达性TLV（IPv6 Reachability TLV）

  类型值为236（0xEC），通过定义路由信息前缀、度量值等信息来说明网络的可达性。IPv6 IS-IS中的IPv6可达性TLV对应于IS-IS中的普通可达性TLV和扩展可达性TLV，格式如图 21 所示。

  图21 IPv6可达性TLV



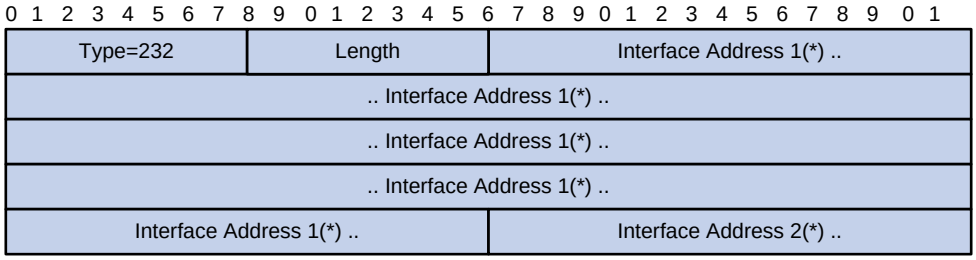
  主要字段的解释如下：

  - Type：取值为236，表示该TLV是IPv6可达性TLV。
  - Length：TLV长度。
  - Metric：度量值，使用扩展的Metric值，取值范围为0～4261412864。度量值大于4261412864的IPv6可达性信息都被忽略掉。
  - U：up/down状态标志位，用来防止路由环路。当某条路由从Level-2路由器传播到Level-1路由器时，这个位被置为1，从而保证了该路由不会被回环。
  - X：外部路由引入标识，取值1表示该路由是从其它协议引入的。
  - S：当TLV中不携带Sub-TLV时，S位置“0”；当S位置“1”时，表示IPv6前缀后面跟随Sub-TLV信息。
  - Reserve：保留位。
  - Prefix Length：IPv6路由的前缀长度。
  - Prefix：该路由器可以到达的IPv6路由前缀。
  - Sub-TLV Length/Sub-TLVs：Sub-TLV字段长度以及Sub-TLVs字段，该选项用于以后扩展用。

- IPv6接口地址TLV

  类型值为232（0xE8），它对应于IPv4中的IP Interface Address TLV，只不过把原来的32比特的IPv4地址改为128比特的IPv6地址。IPv6接口地址TLV对应于IS-IS中的IPv4接口地址TLV，格式如图 22 所示。

图22 IPv6 接口地址 TLV



*: if present

主要字段的解释如下：

- Type：取值为 232，表示该 TLV 类型是 IPv6 接口地址 TLV。
- Length：TLV 长度。
- Interface Address：使能 IPv6 IS-IS 功能接口的 IPv6 地址，Hello 报文中接口 IPv6 地址 TLV 填入的是接口的 IPv6 链路本地地址，LSP 报文中填入的是接口的非 IPv6 链路本地地址，即接口的 IPv6 全球单播地址。

2. IPv6 IS-IS 邻接关系

IS-IS 使用 Hello 报文来发现同一条链路上的邻居路由器并建立邻接关系，当邻接关系建立完毕后，将继续周期性地发送 Hello 报文来维持邻接关系。为了支持 IPv6 路由，建立 IPv6 邻接关系，IPv6 IS-IS 对 Hello 报文进行了扩充：

- NLPID 是标识 IS-IS 支持何种网络层协议的一个 8 比特字段，IPv6 IS-IS 对应的 NLPID 值为 142（0x8E）。如果设备支持 IPv6 IS-IS 功能，那么它必须在 Hello 报文中携带该值向邻居通告其支持 IPv6。
- 在 Hello 报文中添加 IPv6 接口地址 TLV，Interface Address 字段填入使能了 IPv6 IS-IS 功能接口的 IPv6 链路本地地址。

3.3.4 IPv6 BGP（即 BGP4+）

IPv6 BGP 利用 BGP 的多协议扩展属性，来实现在 IPv6 网络中跨自治系统传播 IPv6 路由。

BGP-4 中与 IPv4 网络层协议相关的信息由 Update 消息携带，这些信息是：NLRI 和 NEXT_HOP 属性等路径属性。

为实现对 IPv6 的支持，IPv6 BGP 在 NLRI 和 NEXT_HOP 属性基础上进行了扩展：

- 引入两个新的路径属性 MP_REACH_NLRI 和 MP_UNREACH_NLRI，用以代替 BGP-4 的 NLRI 字段，以提供对 IPv6 地址前缀的 BGP 路由的支持。
- 下一跳信息新增对 IPv6 地址的支持，下一跳信息中不仅可以携带全球单播 IPv6 地址，还可以携带链路本地地址。IPv6 BGP 的下一跳信息通过 MP_REACH_NLRI 属性携带，而不是在 NEXT_HOP 属性中携带。

在尚未完全演进的 IPv4/IPv6 混合网络中，IPv6 BGP 提供了通过 BGP IPvX 会话承载 IPv6 的能力，使得设备可以在 IPv4 会话上交互 IPv6 BGP 路由，亦可以在 IPv6 会话上交互 IPv4 BGP 路由，为 IPv4/IPv6 网络提供了扩展支持 IPv6/IPv4 流量转发的能力。

3.3.5 IPv4 和 IPv6 路由协议异同点总结

IPv4 和 IPv6 路由协议的主要异同点如所示。

表5 IPv4 和 IPv6 路由协议的主要异同点

协议类型	相同点	主要差异点
RIP/RIPng	路由计算思路、基本工作机制相同	• 报文格式的差异（组播地址、UDP 端口、协议报文格式） • 路由下一跳处理的差异 • RIPng 的安全控制由 IPv6 报头实现
OSPFv2/OSPFv3	路由计算思路、基本工作机制相同	• OSPFv3 修改了 LSA 的种类和格式，使其支持发布 IPv6 路由信息 • OSPFv3 修改部分协议流程，使其独立于网络协议，大大提高了可扩展性 • OSPFv3 支持处理未知类型 LSA，提高了协议对未来扩展的适应性
IS-IS/IPv6 IS-IS	协议架构相同	• IPv6 IS-IS 在 Hello 报文中新定义了支持 IPv6 的网络层协议标识符 NLPID（类型值为 142） • IPv6 IS-IS 新增 IPv6 接口地址 TLV 和 IPv6 可达性 TLV
BGP-4/IPv6 BGP	协议架构相同	• IPv6 BGP 扩展 Open 消息，使其支持 IPv6 能力协商 • IPv6 BGP 扩展了支持 IPv6 地址的 MP_REACH_NLRI、MP_UNREACH_NLRI 和 Nexthop 属性

3.4 双栈策略路由

与按照报文目的地址查找路由表进行转发的路由协议不同，策略路由是一种依据用户制定的策略进行转发的机制。策略路由可以对于满足一定条件（例如 ACL 规则）的报文，执行指定的操作（设置报文的下一跳、出接口、缺省下一跳和缺省出接口等）。

双栈策略路由与单栈策略路由（包括 IPv4 策略路由和 IPv6 策略路由）在报文的转发流程上基本相同，主要区别在于单栈策略路由只能处理 IPv4 或 IPv6 一种报文，而双栈策略路由支持同时处理 IPv4 和 IPv6 两种报文。

在双协议栈节点使用双栈策略路由可以减少配置的复杂度，同时节省一定的驱动资源。

3.5 IPv6组播

3.5.1 IPv6 组播简介

组播是指在 IP 网络中将数据包以尽力传送的形式发送到某个确定的节点集合（即组播组），其基本思想是：源主机（即组播源）只发送一份数据，其目的地址为组播组地址；组播组中的所有接收者都可收到同样的数据拷贝，并且只有组播组内的主机可以接收该数据，其它主机无法接收。

组播技术有效地解决了单点发送、多点接收的问题，实现了 IP 网络中点到多点的高效数据传送，能够大量节约网络带宽、降低网络负载。作为一种与单播和广播并列的通信方式，组播的意义不仅在于此。更重要的是，可以利用网络的组播特性方便地提供一些新的增值业务，如在线直播、网络电视、远程教育、远程医疗、网络电台、实时视频会议等互联网的信息服务领域。

IPv6 组播与 IPv4 组播的最大不同在于 IPv6 组播地址机制的极大丰富，而其它诸如组成员管理、组播报文转发以及组播路由建立等与 IPv4 组播基本相同。因此，本文将重点介绍组播地址对 IPv6 的支持情况；对于 IPv6 组播协议，只对其与 IPv4 组播协议的异同进行大致的介绍。

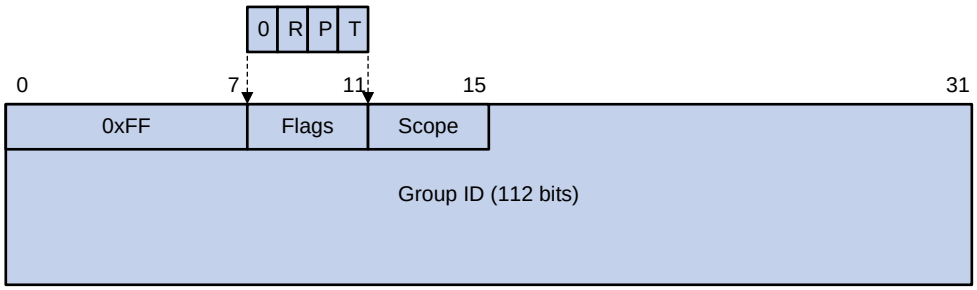
3.5.2 IPv6 组播地址

在介绍 IPv6 组播地址之前，先简单回顾一下 IPv6 的地址结构：IPv6 地址的长度为 128 比特，每个 IPv6 地址被分为 8 组，每组的 16 比特用 4 个十六进制数来表示，组和组之间用冒号隔开，例如：FEDC:BA98:7654:3210:FEDC:BA98:7654:3210。

1. IPv6 组播地址格式

IPv6 组播地址用来标识一组接口，通常这些接口属于不同的节点。一个节点可能属于 0 到多个组播组。发往组播地址的报文被组播地址标识的所有接口接收。

图23 IPv6 组播地址格式



如图 23 所示，IPv6 组播地址中各字段的含义如下：

- 0xFF：最高 8 比特为 11111111，标识此地址为 IPv6 组播地址。
- Flags：4 比特，该字段中各位的取值及含义如表 6 所示。

表6 Flags 字段各位的取值及含义

位	取值及含义
0位	保留位，必须取0
R位	• 取 0 表示非内嵌 RP 的 IPv6 组播地址 • 取 1 表示内嵌 RP 的 IPv6 组播地址（此时 P、T 位也必须置 1）
P位	• 取 0 表示非基于单播前缀的 IPv6 组播地址 • 取 1 表示基于单播前缀的 IPv6 组播地址（此时 T 位也必须置 1）
T位	• 取 0 表示由 IANA 永久分配的 IPv6 组播地址 • 取 1 表示非永久分配的 IPv6 组播地址

- Scope：4 比特。用来标识该 IPv6 组播组的应用范围，其取值及含义如表 7 所示。

表7 Scope 字段的取值及其含义

取值	含义
0、F	保留（Reserved）
1	接口本地范围（Interface-Local Scope）
2	链路本地范围（Link-Local Scope）
3	子网本地范围（Subnet-Local Scope）
4	管理本地范围（Admin-Local Scope）
5	站点本地范围（Site-Local Scope）
6、7、9～D	未分配（Unassigned）
8	机构本地范围（Organization-Local Scope）
E	全球范围（Global Scope）

- Group ID：112 比特，IPv6 组播组标识号。用来在由 Scope 字段所指定的范围内唯一标识 IPv6 组播组，该标识可能是永久分配的或临时的，这由 Flags 字段的 T 位决定。

2. 预留的 IPv6 组播地址

根据 RFC 4291，目前已被预留的 IPv6 组播地址如表 8 所示。

表8 预留的 IPv6 组播地址列表

名称	地址	说明
保留组播地址	FF0X::	不能分配给任何组播组
所有节点组播地址	• FF01::1（节点本地） • FF02::1（链路本地）	-
所有路由器组播地址	• FF01::2（节点本地） • FF02::2（链路本地） • FF05::2（站点本地）	-
被请求节点组播地址	FF02::1:FFXX:XXXX	在被请求节点单播或任播IPv6地址的低24位前增加地址前缀FF02::1:FF00::/104而得，如4037::01:800:200E:8C6C对应于FF02::1:FF0E:8C6C

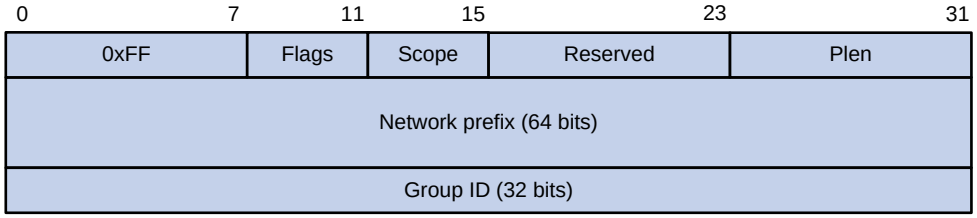
说明

表 8 中的 X 代表 0～F 的任意一个十六进制数。

3. 基于单播前缀的 IPv6 组播地址

RFC 3306 中规定了一种动态分配 IPv6 组播地址的方式——基于单播前缀的 IPv6 组播地址。这种 IPv6 组播地址中包含了其组播源网络的单播地址前缀，通过这种方式分配全局唯一的组播地址。

图24 基于单播前缀的 IPv6 组播地址格式



基于单播前缀的 IPv6 组播地址的格式如图 24 所示，其中各字段的含义如下：

- Flags：R 位置 0，P、T 位则分别置 1，表示基于单播前缀的组播地址。
- Scope：如 1. 图 23 表 7 所示。
- Reserved：8 比特。保留字段，必须为 0。
- Plen：8 比特。表示网络前缀的有效长度（单位为比特）。
- Network prefix：64 比特。表示该组播地址所属子网的单播前缀，有效长度由 Plen 字段指定。
- Group ID：32 比特。表示 IPv6 组播组标识号。

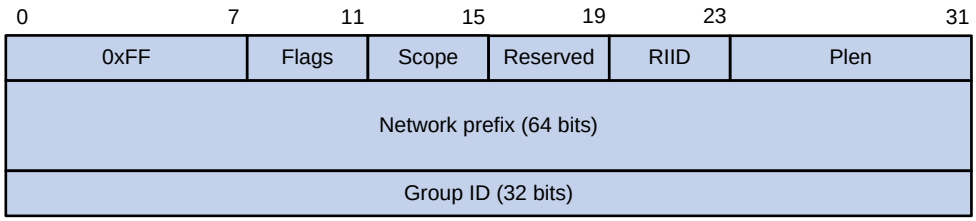
例如：单播前缀为 3FFE:FFFF:1::/48 的网络分配基于单播前缀的组播地址为 FF3X:30:3FFE:FFFF:1::/96（X 表示任意合法的 Scope）。

4. 内嵌 RP 地址的 IPv6 组播地址

- 地址格式

  嵌入式 RP（Rendezvous Point，汇集点）是 IPv6 PIM 中特有的 RP 发现机制，该机制使用内嵌 RP 地址的 IPv6 组播地址，使得组播路由器可以直接从该地址中解析出 RP 的地址。

  图25 内嵌 RP 地址的 IPv6 组播地址格式



  如图 25 所示，内嵌 RP 地址的 IPv6 组播地址使用基于单播前缀的 IPv6 组播地址格式，其中各字段的含义如下：

  - Flags：R、P 和 T 位均置 1，表示内嵌 RP 地址的组播地址。
  - Scope：如 1. 图 23 表 7 所示。
  - Reserved：4 比特。保留字段，必须为 0。
  - RIID：4 比特。表示 RP 地址的接口 ID。
  - Plen：8 比特。表示 RP 地址前缀的有效长度（单位为比特）。
  - Network prefix：64 比特。表示 RP 地址前缀，有效长度由 Plen 字段指定。
  - Group ID：32 比特。表示 IPv6 组播组标识号。

- 计算规则

  内嵌于 IPv6 组播地址中的 RP 地址的计算规则如下：

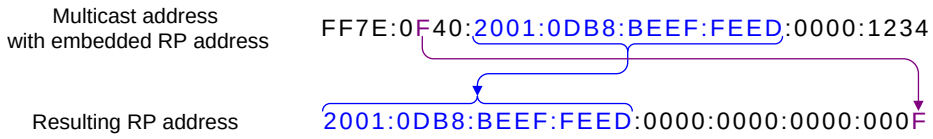
  a. 先将 IPv6 组播地址 Network prefix 字段的前 Plen 位作为 RP 地址的网络前缀。

  b. 再将 IPv6 组播地址 RIID 字段填充到 RP 地址的最低 4 位。

  c. 最后，将 RP 地址的所有剩余位补 0。

  例如：对于 IPv6 组播地址 FF7E:F40:2001:DB8:BEEF:FEED::1234，内嵌于其中的 RP 地址的前缀为 Network prefix 字段的前 Plen（这里为 0x40 = 64 bits）位，最低 4 位为 RIID（0xF），其余位均为 0，如图 26 所示。

  图26 嵌入式 RP 计算举例



- 应用举例

  假设网络管理员想在 2001:DB8:BEEF:FEED::/64 网段中设置 RP，则内嵌 RP 地址的 IPv6 组播地址为 FF7X:Y40:2001:DB8:BEEF:FEED::/96，可分配 32 比特的 Group ID，内嵌于其中的 RP 地址为 2001:DB8:BEEF:FEED::Y/64。

  如果网络管理员想在 IPv6 组播地址中保留更多可分配的 Group ID，可以选择更短的 RP 地址前缀：譬如取 Plen = 0x20 = 32 bits，则此时内嵌 RP 地址的 IPv6 组播地址为 FF7X:Y20:2001:DB8::/64，可分配 64 比特的 Group ID，内嵌于其中的 RP 地址为 2001:DB8::Y/32。

  说明

  X 表示任意合法的 Scope，Y 代表 1～F 的任意一个十六进制数。

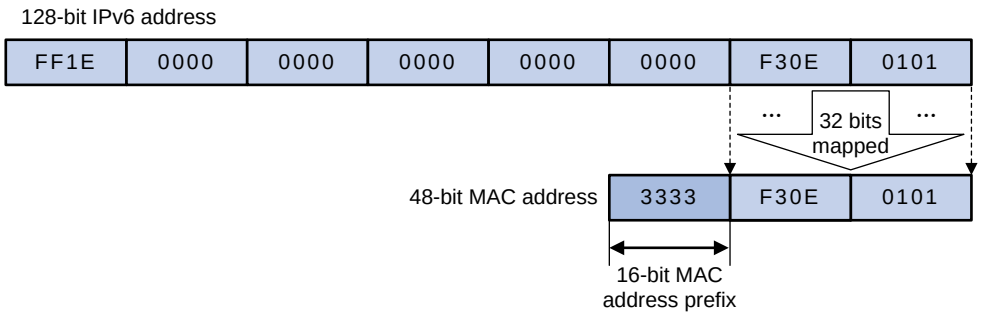
5. IPv6 SSM 组播地址

IPv6 SSM（Source-Specific Multicast，指定信源组播）组播地址也使用基于单播前缀的 IPv6 组播地址格式，其中的 Plen 字段和 Network prefix 字段均取 0。IPv6 SSM 组播地址范围为 FF3X::/32（X 表示任意合法的 Scope）。

3.5.3 IPv6 组播 MAC 地址

IPv6 组播 MAC 地址以 0x3333 开头，低 32 位为 IPv6 组播地址的低 32 位，最终形成 48 比特的组播 MAC 地址。如图 27 所示，IPv6 组播地址 FF1E::F30E:101 所对应的组播 MAC 地址为 33-33-F3-0E-01-01。

图27 IPv6 组播地址的 MAC 地址映射举例



3.5.4 IPv6 组播协议

IPv6 支持的组播协议包括 MLD（Multicast Listener Discovery Protocol，组播侦听者发现协议）、MLD Snooping（Multicast Listener Discovery Snooping，组播侦听者发现协议窥探）、IPv6 PIM（IPv6 Protocol Independent Multicast，IPv6 协议无关组播）和 IPv6 MBGP（IPv6 Multicast BGP，IPv6 组播 BGP）等。

1. 组播组管理协议

MLD 源自 IGMP（Internet Group Management Protocol，互联网组管理协议），MLD 有两个版本：MLDv1 源自 IGMPv2，MLDv2 源自 IGMPv3。

与 IGMP 采用 IP 协议号为 2 的报文类型不同，MLD 采用 ICMPv6（IP 协议号为 58）的报文类型，包括 MLD 查询报文（类型值 130）、MLDv1 报告报文（类型值 131）、MLDv1 离开报文（类型值 132）和 MLDv2 报告报文（类型值 143）。MLD 协议与 IGMP 协议除报文格式不同外，协议行为完全相同。

2. 组播路由协议

IPv6 PIM 与 PIM 除报文中 IP 地址结构不同外，其它协议行为基本相同，IPv6 PIM 也支持如下四种模式：

- IPv6 PIM-DM（IPv6 Protocol Independent Multicast-Dense Mode，IPv6 协议无关组播—密集模式）
- IPv6 PIM-SM（IPv6 Protocol Independent Multicast-Sparse Mode，IPv6 协议无关组播—稀疏模式）
- IPv6 PIM-SSM（IPv6 Protocol Independent Multicast Source-Specific Multicast，IPv6 协议无关组播－指定源组播）
- IPv6 BIDIR-PIM（IPv6 Bidirectional Protocol Independent Multicast，IPv6 双向协议无关组播，简称 IPv6 双向 PIM）

IPv6 PIM 发送链路本地范围的协议报文（包括 PIM Hello、Join-Prune、Assert、Bootstrap、Graft、Graft-Ack 和 State-refresh 报文）时，报文的源 IPv6 地址使用发送接口的链路本地地址；IPv6 PIM 发送全球范围的协议报文（包括 Register、Register-Stop 和 C-RP Advertisement 报文）时，报文的源 IPv6 地址使用发送接口的全球单播地址。

IPv6 组播并不支持 MSDP 协议，如果需要接收来自其它 IPv6 PIM 域的组播数据，有以下两种实现方式：

- 通过其它方式（譬如广告等）直接获取其它 IPv6 PIM 域内的组播源地址，使用 IPv6 PIM-SSM 发起指定源组的加入。
- 使用嵌入式 RP 机制，通过嵌入 RP 地址的 IPv6 组播地址来获取其它 IPv6 PIM 域内的 RP 地址，向其它域内的 RP 发起组加入。

对于域间 IPv6 组播路由信息的传递，则可以使用 IPv6 的 MBGP 协议，其与 IPv4 的 MBGP 协议也基本相同。

3. 二层组播协议

- MLD Snooping

  MLD Snooping 与 IGMP Snooping 协议基本相同。
- IPv6 PIM Snooping

  IPv6 PIM Snooping 与 PIM Snooping 协议基本相同。
- 组播 VLAN

  组播 VLAN，对于 IPv4 组播和 IPv6 组播，处理原理相同。

3.6 网络安全

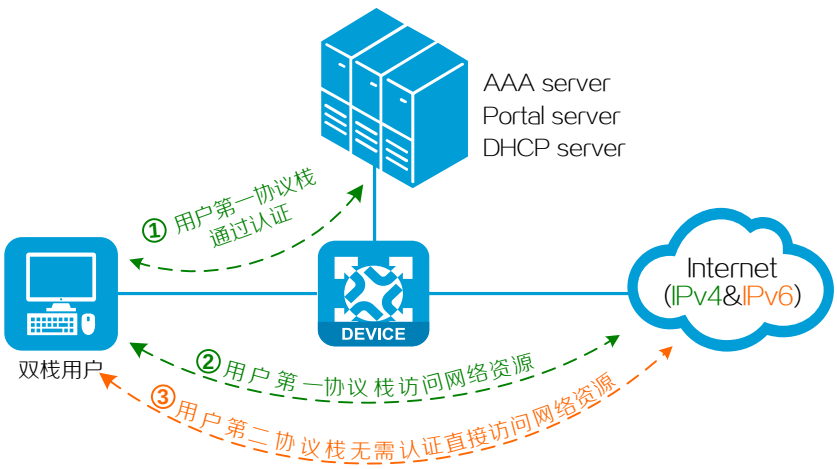
3.6.1 一次认证双栈放行

1. 概述

在 IPv4 网络完全过渡到 IPv6 网络之前，若用户主机同时支持 IPv4 和 IPv6 两种协议，可能会产生两种地址协议类型的流量分别触发对应协议栈的 IPoE Web 认证或 Portal 认证。如果用户只通过 IPv4 或 IPv6 其中一种认证，就只能访问对应协议栈的网络资源。如果用户需要进行两次不同协议类型的认证，又会增加用户上网操作的复杂度。

IPoE Web 认证或 Portal 认证支持双栈技术可以很好地解决上述问题，它可以实现一次认证双栈放行，即当双栈用户通过任意一个协议栈（IPv4 或 IPv6）流量触发认证并成功上线后，它的另外一个协议栈流量无需认证即被放行。

图28 一次认证双栈放行示意图



该技术为 IPv6 网络带来如下技术价值：

- 对于用户，两个协议栈上线只需要完成一次认证过程，提升使用体验。
- 对于服务器，用户双栈上线只需要进行一次认证，减轻了 AAA 服务器和 Portal 服务器的认证压力。
- 对于管理员，同一用户的 IPv4 和 IPv6 协议栈作为一个双栈用户进行处理，降低了网络管理和维护的复杂度。

2. IPoE Web 认证支持双栈技术

在 IPoE Web 双栈认证组网中，双栈用户上线的基本过程如下：

(1) 用户通过第一协议栈（如 IPv4）上线时，在认证页面中输入用户名和密码，认证通过后可访问相应协议栈的网络资源。设备上记录该用户的 MAC 地址、用户名和认证状态等信息。

(2) 用户通过第二协议栈（如 IPv6）上线时，设备根据用户的 MAC 地址判断该用户的另一协议栈是否已上线。如上线，则视为同一用户，第二协议栈无需再次认证，直接放行。

根据用户两个协议栈上线方式的不同，IPoE Web 双栈认证支持如下三种典型应用场景：

- 动态双栈用户上线：双栈用户可以通过未知源 IPv4 报文、未知源 IPv6 报文、DHCPv4 报文、DHCPv6 报文或 ND RS 报文触发动态上线。该方式多用于移动终端非固定 IP 的场景。例如，学生通过移动智能终端接入校园网。
- 静态双栈用户上线：双栈用户可以通过 IPv4 报文、IPv6 报文、ARP 报文 NS 报文或 NA 报文触发静态上线。该方式多用于终端 IP 地址固定的场景。例如，学生在宿舍通过固定网口接入校园网。
- 混合双栈用户上线：双栈用户的一个协议栈采用静态方式上线，另一个协议栈采用动态方式上线。该方式多用于网络中同时存在固定 IP 和非固定 IP 终端的场景。例如，某高校的原有的 IPv4 网络用户采用静态 IPv4 地址方式，学校对现网进行 IPv6 改造升级后，使得原有 IPv4 用户可以接入 IPv6 网络，同时希望用户的 IPv6 地址通过 DHCPv6 动态分配，即采用静态 IPv4+动态 IPv6 的混合地址分配方式。

3. Portal 认证支持双栈技术

在 Portal 双栈认证组网中，管理员根据现网的实际需求，在使能了 Portal 认证的接口上开启 Portal 支持双栈认证功能后，该接口上的用户只需要通过 IPv4 Portal 或 IPv6 Portal 认证中的任何一种，就可以访问 IPv4 和 IPv6 两种协议栈对应的网络资源。

Portal 双栈用户上线的基本过程如下：

(1) 双栈用户的第一协议栈（IPv4 或 IPv6）报文触发 Portal 认证后，用户在 Portal Web 认证页面中输入用户名和密码，之后若通过 IPv4 或 IPv6 Portal 认证，则可访问对应协议栈的网络资源。

(2) 设备将通过 IPv4 Portal 认证或 IPv6 Portal 认证的用户 MAC 地址和 IP 地址记录在 Portal 用户表项中。

(3) 设备收到该用户的第二协议栈（IPv6 或 IPv4）任意报文时，如果报文中的源 MAC 地址与记录在 Portal 用户表项中的 MAC 地址相同，则允许其访问对应协议栈的网络资源，不需要再次进行认证。

 说明

仅当接口上同时开启了直接认证方式的 IPv4 和 IPv6 Portal 认证功能，Portal 支持双栈认证功能才会生效。

3.6.2 SAVI&SAVA&SMA

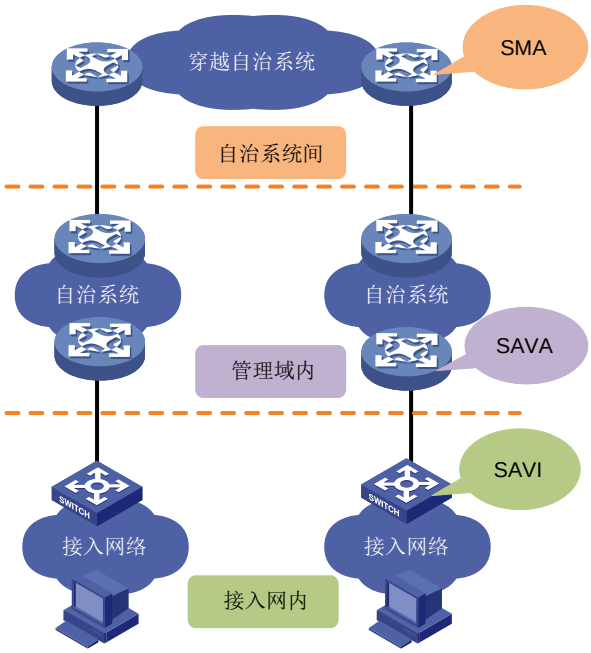
1. 概述

真实 IPv6 源地址验证是指，通过对报文的 IPv6 源地址进行验证，丢弃伪造 IPv6 源地址的报文，提升 IPv6 网络的安全性。SAVI&SAVA 均属于真实 IPv6 源地址验证技术，分别部署在不同的网络位置，能够满足不同粒度的安全需求。

根据在网络中部署位置的不同，真实 IPv6 源地址验证功能分为如下三种类型：

- SAVI（Source Address Validation Improvement，源地址有效性验证）：部署在接入网，在接入层面提供主机粒度的源地址验证，保证主机只使用合法分配的 IPv6 地址。
- SAVA（Source Address Validation Architecture，源地址验证架构）：部署在骨干网连接接入网的边界设备上，在管理域内提供 IPv6 前缀粒度的保护能力，以保护核心设备不被仿冒源地址的非法主机攻击。
- SMA（State Machine based Anti-spoofing，基于状态机的伪造源地址检查）：部署在 AS 间，在 AS 域间提供 AS 粒度的源地址验证能力，以保护本 AS 内的主机和服务器不被仿冒源地址的非法主机攻击。

图29 SAVI&SAVA&SMA 在网络中部署位置示意图

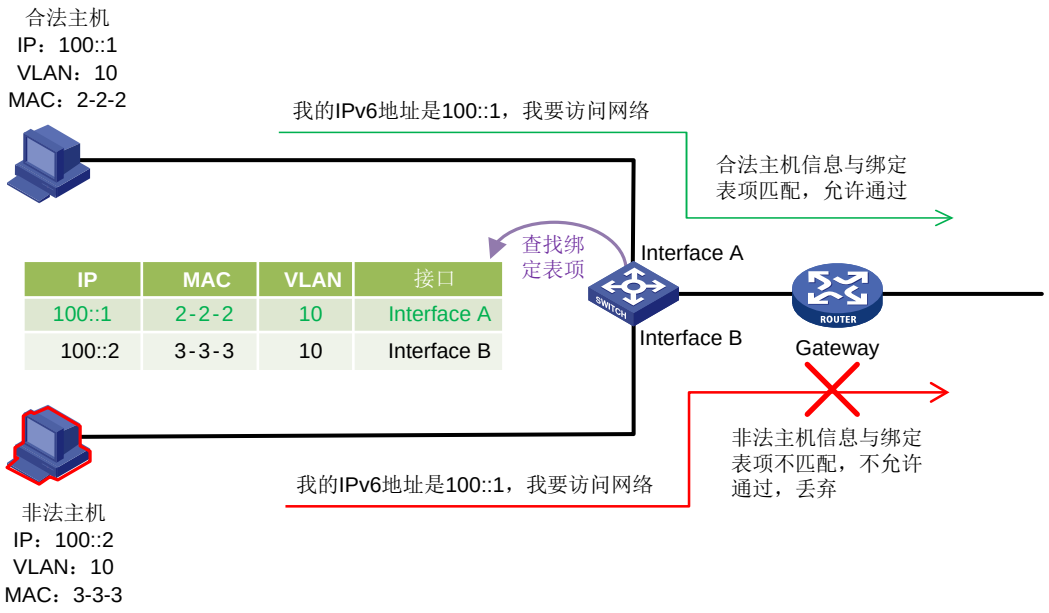


2. SAVI

为了防止 IPv6 源地址非法的 DHCPv6 协议报文、ND 协议报文和 IPv6 数据报文形成攻击，可以在设备上开启 SAVI 功能。设备在其它安全功能的配合下，生成绑定表项，并根据该绑定表项对报文 IPv6 源地址进行检查。如果报文信息与某绑定表项匹配，则认为该报文为合法报文，正常转发；否则将该报文丢弃。

与 SAVI 配合使用的安全功能包括 DHCPv6 Snooping、ND Snooping 和 IP Source Guard 中 IPv6 静态绑定表项功能。

图30 SAVI 原理图



3. SAVA

SAVA 是一种根据设备的路由信息检查攻击报文的技术，用来防范基于 IPv6 源地址欺骗的攻击，主要部署在与接入网相连的骨干网内边界设备上。在设备的接入网侧接口上开启 SAVA 功能后，设备会为该接入网络中的所有的网络前缀生成 SAVA 表项。该接口收到 IPv6 报文后，如果存在报文 IPv6 源地址对应的 SAVA 表项，则认为该 IPv6 源地址合法，转发该报文；否则，表示报文 IPv6 源地址不应该存在于接入网络中，报文非法，被丢弃。

以边界设备 B 为例，SAVA 表项生成过程如图 31 所示，分为如下几个步骤：

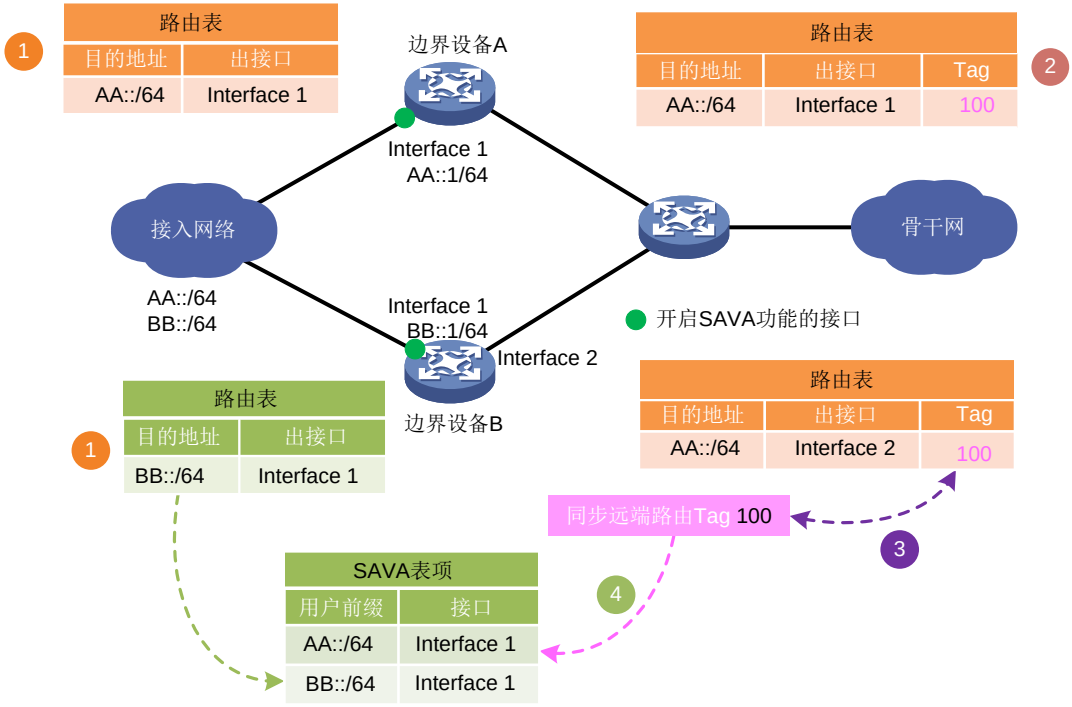
(1) 边界设备 A 和 B 分别从本地学习的、到达接入网络的路由信息中获取用户前缀，这些路由信息包括与接入网络相连的直连路由、静态路由和动态路由。本例中以静态路由为例来说明。

(2) 边界设备 A 为本地学习的、到达接入网络的路由信息打上特定的 Tag，并将此路由信息引入骨干网的动态路由协议中。

(3) 边界设备 B 通过动态路由协议学习到设备 A 发布的带有 Tag 的路由信息。如果路由信息中的 Tag 值与边界设备 B 上配置的同步远端路由条目的 Tag 值相同，则边界设备 B 从该路由信息中获取边界设备 A 学习到的合法用户前缀信息，用于生成 SAVA 表项。

(4) 边界设备 B 将根据本地路由和远端同步路由获取到的所有的合法用户前缀信息来生成与该接口绑定的 SAVA 表项。SAVA 表项信息包含合法用户前缀、前缀长度和绑定的接口。

图31 SAVA 表项生成过程



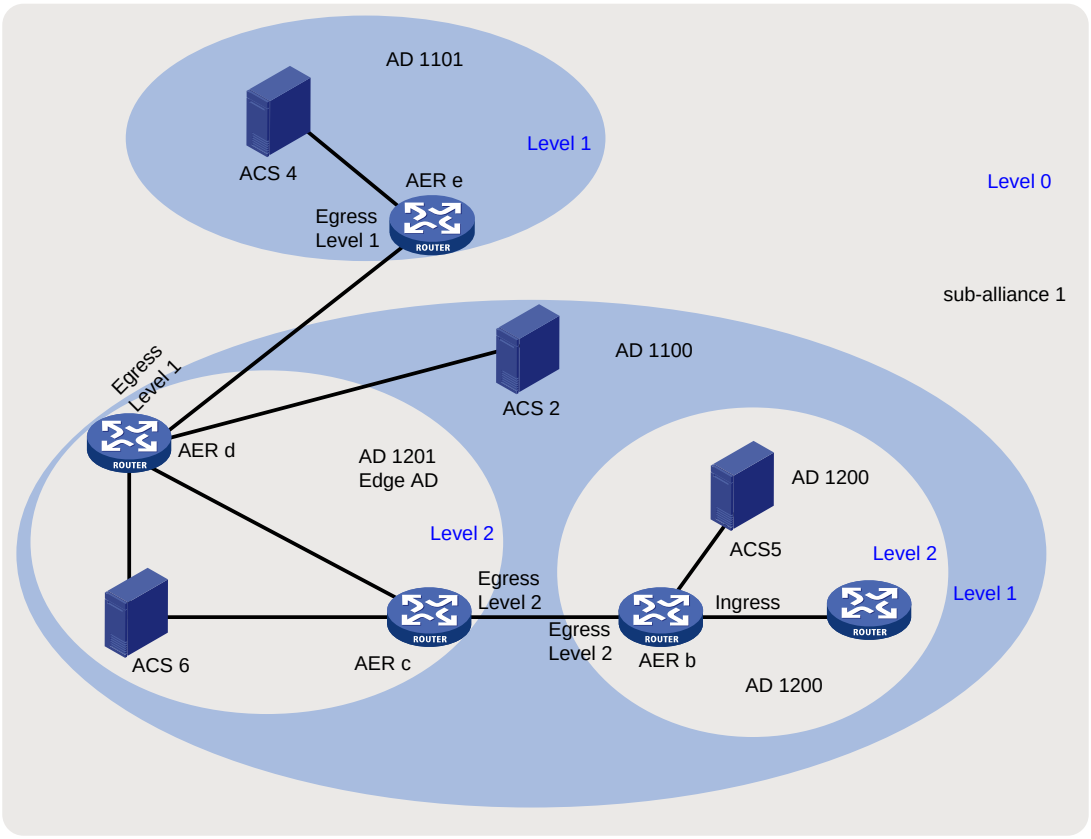
4. SMA

SMA（State Machine based Anti-spoofing，基于状态机的伪造源地址检查）是一种 IPv6 自治系统间端到端的源地址验证方案，用来防止伪造源 IPv6 地址的攻击。

(1) 体系结构

SMA 体系结构主要由 ACS（AS Control Server，AS 控制服务器）和 AER（AS Edge Router，AS 边界路由器）构成，如图 32 所示。

图32 SMA 体系结构



- 子信任联盟：彼此信任的一组 AD（Addrees Domain，地址域）组成的集合，通过子信任联盟号来标识，比如上图中的 sub-alliance 1。
- 信任联盟：SMA 体系中所有 AD 的集合。
- AD（Addrees Domain，地址域）：同一个机构下所管理的所有 IP 地址部署的范围，是子信任联盟管理的对象，通过地址域编号来标识，比如，上图中的 AD 1101、AD 1200 和 AD 1201。同一个子联盟内的不同的地址域可以分成不同的地址域层级，最多可以划分为 4 层。比如，上图中的 Level 0、Level 1 和 Level 2。其中，Level 0 为最高地址级别，Level 2 为最低地址级别。例如，首先以县市为单位划分多个一级地址域，再以机构为单位划分多个二级地址域（比如学校、企事业单位），以楼宇或部门为单位划分三级地址域。
  - 边界地址域：当前层级的地址域中与其他层级相连的地址域。比如，上图中的 AD 1201。
  - 非边界地址域：除了边界地址域的其他地址域。

  当一个地址域划分了更低级别的地址域后，原地址域中所有的设备都必须从属于更低级别的地址域中。如上图所示，Level 0 的地址域中划分了低一级别的地址域 Level 1，那么属于 Level 0 的所有设备都必须从属于划分后的 Level 2 地址域。
- ACS（AS Control Server，AS 控制服务器）：每个层级的地址域都需要有相应的 ACS，用于和其它地址域内的 ACS 交互信息，并向本地址域内的 AER 宣告与更新注册信息、前缀信息以及状态机信息。具体来讲，ACS 具有如下功能：

- 与属于相同信任联盟中各子信任联盟的其他 ACS 建立连接，交互各地址域内的 IPv6 地址前缀、状态机等信息。
- 向本地址域 AER 宣告和更新联盟映射关系、地址前缀信息以及标签信息。

○ AER（AS Edge Router，AS 边界路由器）：负责接收 ACS 通告的 IPv6 地址前缀、标签等信息，并在地址域之间转发报文。一个 AER 可以是多个不同层级 ACS 的边界路由器。AER 上的接口分为两类：

- Ingress 接口：连接到本地址域内部未使能 SMA 特性的路由器的接口。
- Egress 接口：连接其他地址域的 AER 的 Egress 接口。

说明

- 目前，设备只能作为 AER。
- 为了提高安全性，ACS 与 ACS 之间、ACS 与 AER 之间的通信均可配置为基于 SSL（Secure Sockets Layer，安全套接字层）的连接。

---

(2) 工作原理

SMA 通过在 AER 上检查报文的源 IPv6 地址和报文标签实现对伪造源 IPv6 地址攻击的防御。AER 接收到报文后，处理过程如图 33 所示。

a. 检查接收报文的接口类型是否为 Ingress。
   - 若接口类型是 Ingress，则进入步骤 b。
   - 若接口类型是 Egress，则进入步骤 c。

b. 检查报文源地址前缀是否属于当前地址域。
   - 若属于当前地址域，则继续按照 IPv6 路由表转发该报文。
   - 若不属于当前地址域，则丢弃报文。

c. 检查报文目的 IPv6 地址是否属于本子联盟：
   - 若不属于本子联盟，则继续按照 IPv6 路由表转发该报文。
   - 若属于本子联盟，则进入步骤 d。

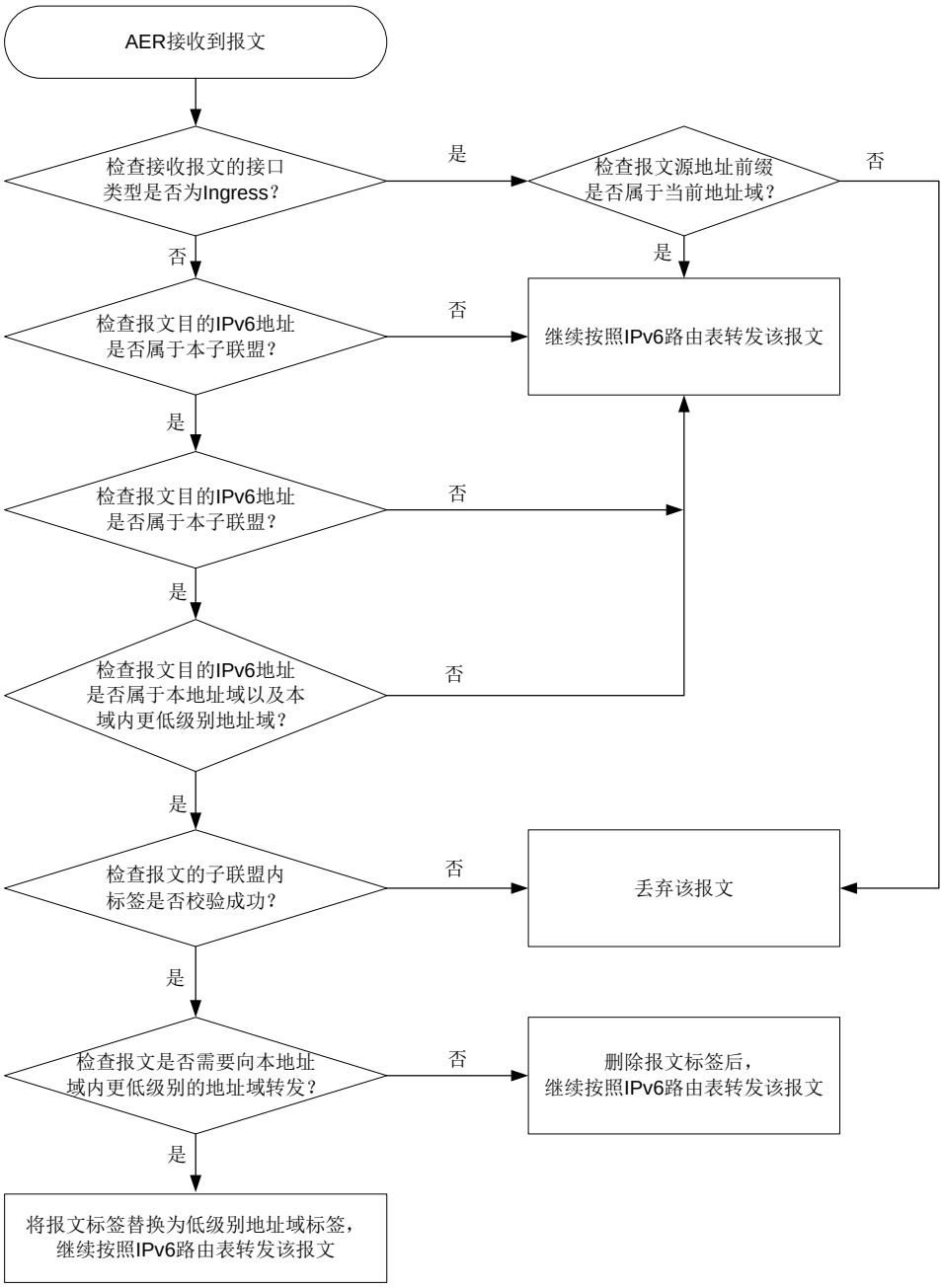
d. 检查报文的目的 IPv6 地址是否属于本地址域以及本域内更低级别地址域：
   - 若不属于，则继续按照 IPv6 路由表转发该报文。
   - 若属于，则进入步骤 e。

e. 校验报文的子联盟内标签。
   - 校验成功，进入步骤 f。
   - 校验失败，丢弃该报文。

f. 检查报文是否需要向本地域内更低级别的地址域转发。
   - 若不需要，删除报文标签后，继续按照 IPv6 路由表转发该报文。
   - 若需要，则将报文标签替换为低级别地址域标签，继续按照 IPv6 路由表转发该报文。

低级别地址域内的 AER 收到报文后，继续按照如上步骤进行处理。

图33 AER 接收报文时的处理过程



当 AER 收到源自本地址域的报文并发往其他地址域时，处理过程如图 34 所示。

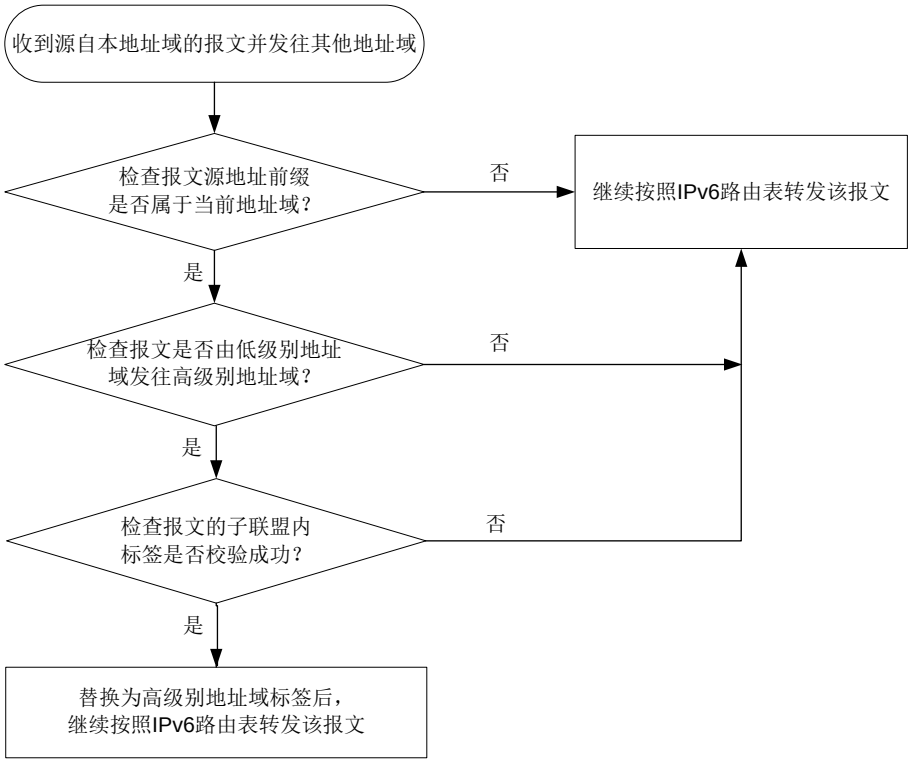
a. 判断报文源地址前缀是否属于当前地址域：
   - 若属于，对报文添加标签后，继续按照 IPv6 路由表转发该报文。
   - 若不属于，则进入步骤 b。

b. 检查报文是否由低级别地址域发往高级别地址域。
   - 若是，则需要校验标签，进入步骤 c。
   - 若不是，则继续按照 IPv6 路由表转发该报文。

c. 校验报文的子联盟内的标签：

– 校验成功，替换为高级别地址域标签后，继续按照 IPv6 路由表转发该报文。
– 校验失败，丢弃该报文。

图34 AER 发送报文时的处理过程



3.6.3 微分段

1. 微分段简介

随着数据中心的不断发展，数据中心网络内部的流量（即东西向流量）在不断增加，数据中心网络流量从以前的南北向流量为主转变为东西向流量为主。网络管理员也需要对东西向流量进行安全防控。如果将数据中心内部东西向流量全部绕行传统的集中式防火墙，很难满足数据中心灵活可扩展部署的要求，防火墙容易称为数据中心性能和扩容的瓶颈。

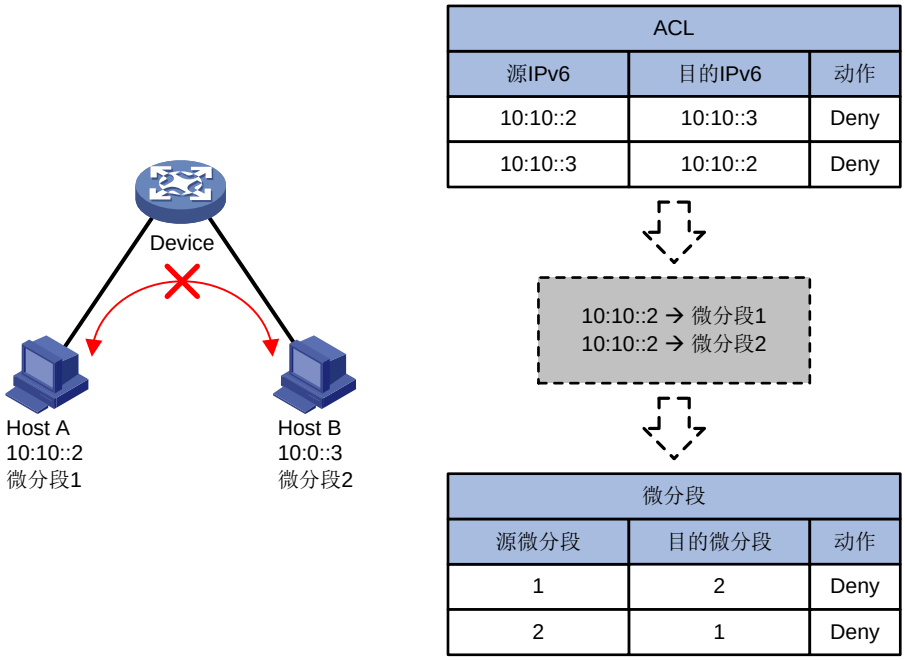
微分段对网络端点（例如数据中心网络中的服务器、虚拟机，或园区网中的各种终端上线用户）进行分组，并部署组间策略，通过组间策略对分属不同组的网络端点之间的通信进行安全管控。这种工作机制决定了它具有管控粒度细和占用 ACL 资源少的优点。

2. 微分段优势

- 分布式安全：微分段方案实现了分布式的安全控制，东西向流量不需要集中转发到防火墙后再进行安全隔离，减少了网络带宽的消耗，可以防止集中式的防火墙成为流量瓶颈。
- 更精细、灵活的安全隔离：传统的 VLAN 或 IP 子网只能实现不同 VLAN 或子网间的隔离，同一 VLAN 或子网内的网络端点无法隔离。同时，当不同子网共用同一个网关设备时，网关设备上保存了到各子网的路由信息。在这种情况下，无法完全实现不同子网内不同网络端点之间的隔离。微分段可基于离散 IP、IP 地址段等进行精细分组，在不同分组之间相应实现更灵活的安全防控。

- 降低 ACL 资源占用：相较于传统的安全管控技术（主要是利用 ACL 的安全管控），微分段技术能够显著降低对 ACL 资源的占用。

  图35 ACL 资源占用示意图



  如图 35 所示，在 IPv6 网络中，如果期望禁止 Host A 和 Host B 互通：

  - 使用 ACL 时，需要匹配报文的 IP 地址。IPv6 地址长度为 128bits，同时匹配源、目的 IP 地址时需要占用 256bits 长度的 ACL 资源。对双向流量同时进行管控时，则需要占用两条长度为 256bits 的 ACL 资源。
  - 使用微分段时，仅需匹配报文所属的微分段 ID。微分段 ID 长度为 16bits，同时匹配源、目的微分段所需的 ACL 资源仅为 32bits。对双向流量同时进行管控时，也仅需两条长度为 32bits 的 ACL 资源。

3. 微分段实现原理

微分段技术中使用到以下概念：

- 微分段是一组网络端点的集合，它通过全局唯一的 ID 来标识。网络管理员可以基于 IP 地址、IP 网段、MAC 地址等对网络端点来划分微分段，以便在网络设备上实现基于微分段 ID 对网络端点进行流量管控。
- GBP（Group Based Policy，组策略）是基于微分段的流量控制策略。通过部署 GBP，可以对属于不同微分段的网络端点之间的通信进行安全管控，相同微分段内的网络端点则可以互访，GBP 不控制相同微分段内的流量。GBP 可以通过报文过滤、QoS 策略（MQC）或策略路由实现。

微分段是一种源端控制策略，即在源端设备上配置微分段功能，实现对流量的安全管控。

微分段功能由三部分组成：

(1) 将网络端点加入微分段。根据应用场景和部署方式的不同，可以通过如下方式将网络端点加入微分段：静态 IP 微分段、静态 AC 微分段、认证授权微分段和路由通告微分段。

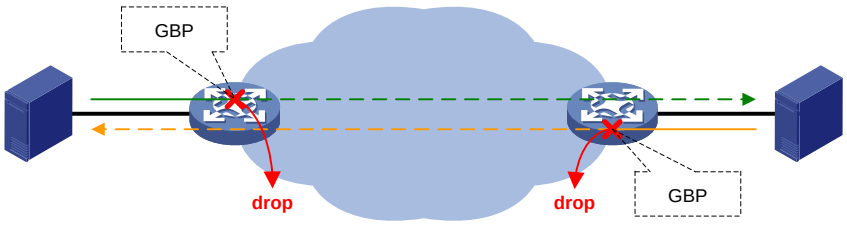
(2) 创建基于微分段 ID 的 ACL。

(3) 使用 GBP，即通过报文过滤、QoS 策略（MQC）或策略路由引用基于微分段 ID 的 ACL，实现对属于不同微分段的网络端点之间的通信进行安全管控。

在源端设备上完成上述配置，源端设备接收到报文后，根据报文所属的微分段 ID，查找匹配的 ACL 规则，再通过 ACL 关联到 GBP。GBP 对命中 ACL 的报文进行流量控制。

综上所述，微分段是生效在报文转发路径中的源端设备上的。当 GBP 判决结果为丢弃时，报文将被直接丢弃，不会再经由中间网络转发至目的端，这就避免了带宽浪费。

图36 源端流量控制示意图



4. 微分段的典型应用举例

微分段可以应用在 EVPN VXLAN 数据中心网络中。通过命令行手工部署或 SDN 控制器自动部署微分段、ACL 和 GBP。

- 对东西向的流量进行管控时，微分段成员的部署方式为：

  在所有 Leaf 上为 VM 的 IP 地址配置全网统一的静态 IP 微分段。

  在 Leaf 1 和 Leaf 2 上都配置：

  - 微分段 1 成员为 192:168:1::0/120。
  - 微分段 2 成员为 192:168:2::0/120。
  - 微分段 3 成员为 192:168:3::0/120。

- 对南北向的流量进行管控时，微分段成员的部署方式为：

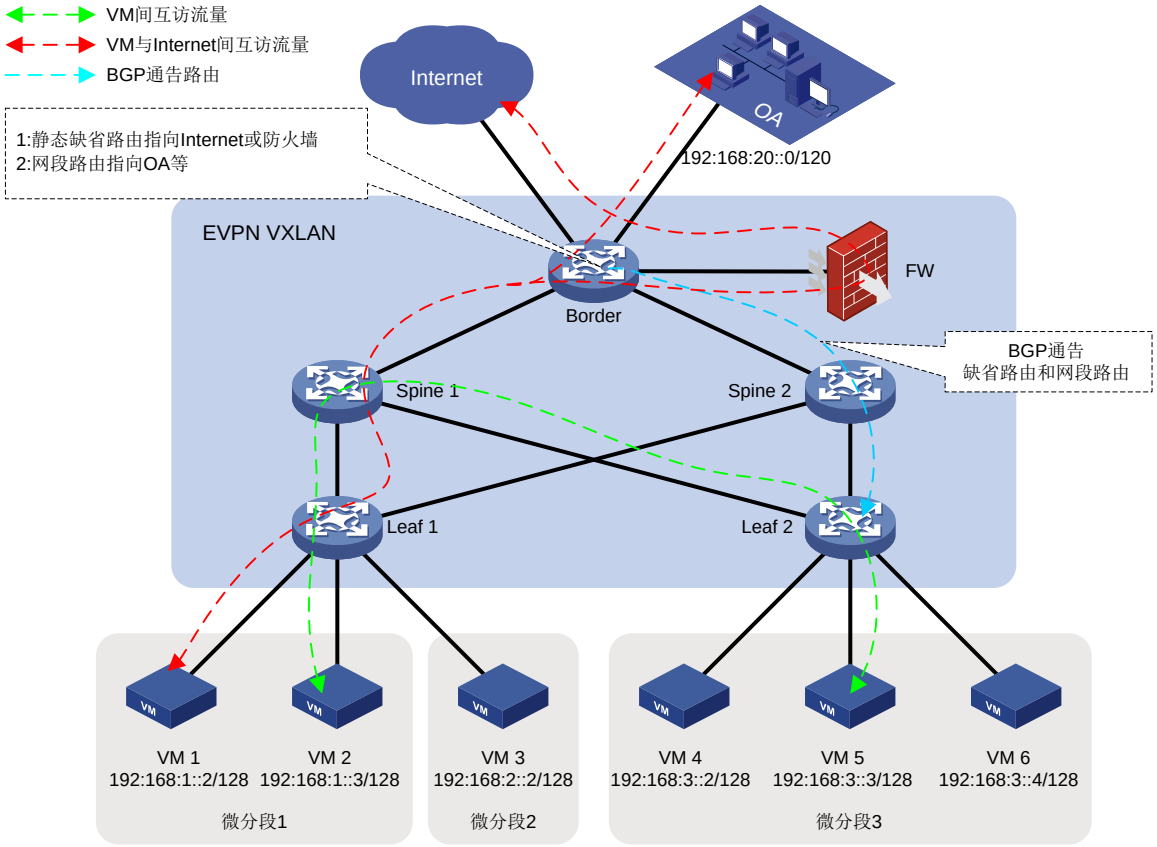
  Border 上存在到达 Internet 和防火墙的静态缺省路由和 OA（办公自动化）网络（即 192:168:20::0/120）的网段路由，通过 BGP 将该路由通告给所有 Leaf，以实现数据中心通过 Border 与外部通信。

  为了实现南北向流量管控，需要在所有 Leaf 上配置全网统一的静态 IP 微分段：

  - 由于 Border 会通告缺省路由，所以在 Leaf 1 和 Leaf 2 上均配置微分段 4，成员为 0::0/0。
  - 由于 Border 会通告网段路由，所以 Leaf 1 和 Leaf 2 上均配置微分段 5，成员为 192:168:20::0/120。

ACL 和 GBP 则按需配置，允许或禁止各微分段间互访的流量通过。

图37 EVPN VXLAN 数据中心网络



3.7 VXLAN/EVPN VXLAN支持IPv6

VXLAN（Virtual eXtensible LAN，可扩展虚拟局域网络）是基于 IP 网络、采用“MAC in UDP”封装形式的二层 VPN 技术。VXLAN 需要手工建立 VXLAN 隧道。

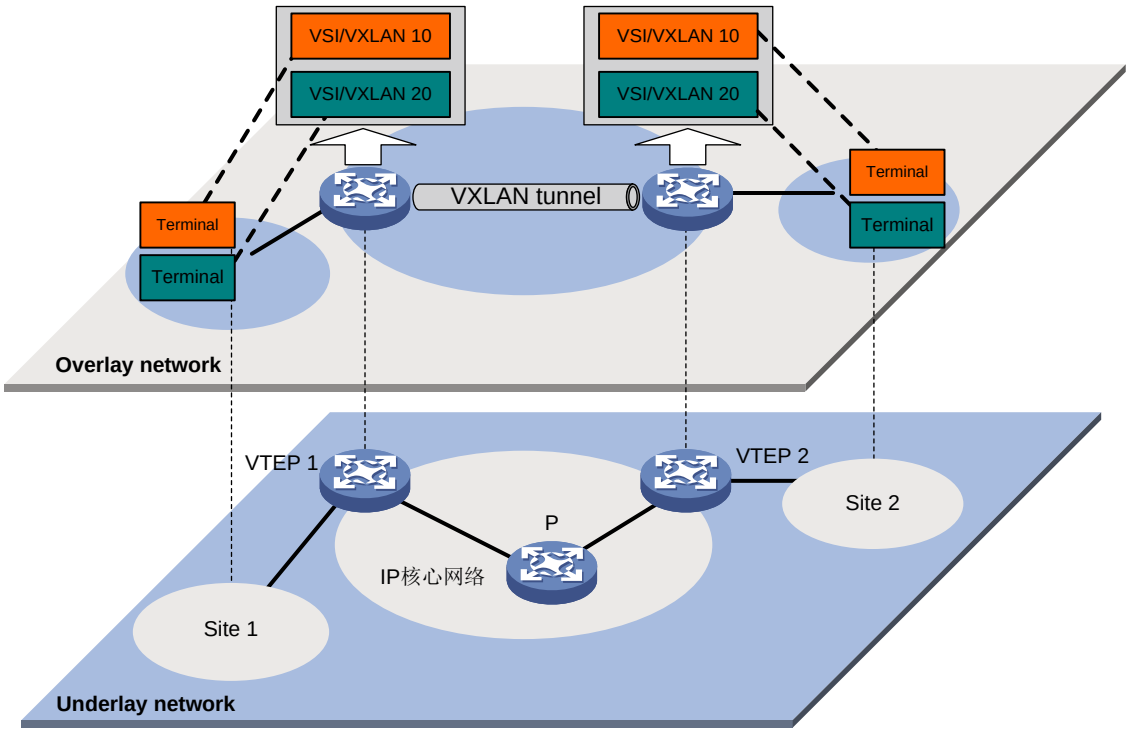
EVPN（Ethernet Virtual Private Network，以太网虚拟专用网络） VXLAN 是一种二层 VPN 技术，控制平面采用 MP-BGP 通告 EVPN 路由信息，数据平面采用 VXLAN 封装方式转发报文。EVPN VXLAN 通过 EVPN 路由自动建立 VXLAN 隧道。

VXLAN/EVPN VXLAN 可以基于已有的服务提供商或企业 IP 网络，为分散的站点网络提供二层互联，实现不同租户的业务隔离。通过网关功能，还可以实现站点网络之间的三层互联。

VXLAN/EVPN VXLAN 技术将已有的三层物理网络作为 Underlay 网络，在其上构建出虚拟的二层网络，即 Overlay 网络。Overlay 网络通过封装技术、利用 Underlay 网络提供的三层转发路径，实现租户二层报文跨越三层网络在不同站点间传递。对于租户来说，Underlay 网络是透明的，同一租户的不同站点就像工作在一个局域网中。

站点网络和 Underlay 网络均可以是 IPv6 网络。

图38 VXLAN/EVPN VXLAN 网络模型示意图



如图 38 所示，VXLAN/EVPN VXLAN 的典型网络模型中包括如下几部分：

- 用户终端（Terminal）：用户终端设备可以是 PC 机、无线终端设备、服务器上创建的 VM（Virtual Machine，虚拟机）等。不同的用户终端可以属于不同的 VXLAN。属于相同 VXLAN 的用户终端处于同一个逻辑二层网络，彼此之间二层互通；属于不同 VXLAN 的用户终端之间二层隔离。VXLAN 通过 VXLAN ID 来标识，VXLAN ID 又称 VNI（VXLAN Network Identifier，VXLAN 网络标识符），其长度为 24 比特。
- VTEP（VXLAN Tunnel End Point，VXLAN 隧道端点）：VXLAN 的边缘设备。VXLAN 的相关处理都在 VTEP 上进行，例如识别以太网数据帧所属的 VXLAN、基于 VXLAN 对数据帧进行二层转发、封装/解封装报文等。
- VXLAN 隧道：两个 VTEP 之间的点到点逻辑隧道。VTEP 为数据帧封装 VXLAN 头、UDP 头和 IP 头后，通过 VXLAN 隧道将封装后的报文转发给远端 VTEP，远端 VTEP 对其进行解封装。
- 核心设备：IP 核心网络中的设备（如图 38 中的 P 设备）。核心设备不参与 VXLAN 处理，仅需要根据封装后报文的目的 IP 地址对报文进行三层转发。
- VSI（Virtual Switch Instance，虚拟交换实例）：VTEP 上为一个 VXLAN 提供二层交换服务的虚拟交换实例。VSI 可以看作 VTEP 上的一台基于 VXLAN 进行二层转发的虚拟交换机，它具有传统以太网交换机的所有功能，包括源 MAC 地址学习、MAC 地址老化、泛洪等。VSI 与 VXLAN 一一对应。
- AC（Attachment Circuit，接入电路）：VTEP 连接本地站点的物理电路或虚拟电路。在 VTEP 上，与 VSI 关联的三层接口或以太网服务实例（service instance）称为 AC。其中，以太网服

务实例在二层以太网接口上创建，它定义了一系列匹配规则，用来匹配从该二层以太网接口上接收到的数据帧。

4 过渡技术

IPv6 网络的部署不是一蹴而就的，在一段时间内 IPv4 网络会与 IPv6 网络共存。过渡技术用来解决 IPv4 网络与 IPv6 网络共存和互通的问题。常用的过渡技术包括双栈、隧道、协议转换（AFT）和 6PE。

4.1 双栈技术

双栈技术是一种最简单直接的过渡机制。双栈技术是指网络中的节点同时支持 IPv4 和 IPv6 两个协议栈，这样的节点称为双协议栈节点。当双协议栈节点配置 IPv4 地址和 IPv6 地址后，就可以在相应接口上转发 IPv4 和 IPv6 报文。当一个上层应用同时支持 IPv4 和 IPv6 协议时，根据协议要求可以选用 TCP 或 UDP 作为传输层的协议，但在选择网络层协议时，它会优先选择 IPv6 协议栈。

双栈技术是所有过渡技术的基础。双栈技术具有如下优点：

- 技术成熟，不必为不同类型的用户单独部署网络配置。
- 开销相对较小，保护用户投资。
- 过渡平滑，通过 IPv6 优选逐步提高 IPv6 流量占比。
- 实现快速业务互访，互通性好，降低了跨协议访问时的地址转换损耗。

双栈技术的缺点为：

- 改造工作量大，需完成整网的 IPv4 和 IPv6 部署，配置管理也较为复杂。
- 设备性能要求高，需考虑设备硬件资源表项共享问题。
- 要求双协议栈节点拥有一个全球唯一的 IPv4 地址和 IPv6 地址，实际上没有解决 IPv4 地址资源匮乏的问题。

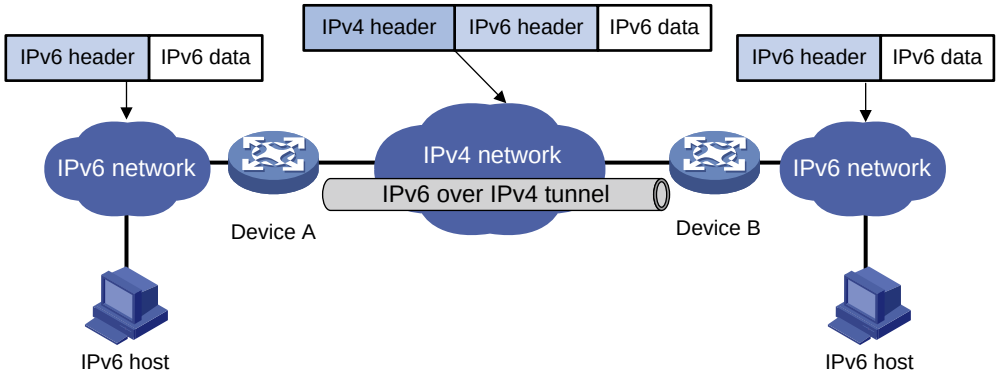
4.2 隧道技术

隧道是一种封装技术，它利用一种网络协议来传输另一种网络协议，即利用一种网络传输协议，将其他协议产生的数据报文封装在它自己的报文中，然后在网络中传输。

IPv6 over IPv4 隧道和 IPv4 over IPv6 隧道可以用于连接 IPv6 或 IPv4 的信息孤岛，解决 IPv4 网络与 IPv6 网络共存问题：

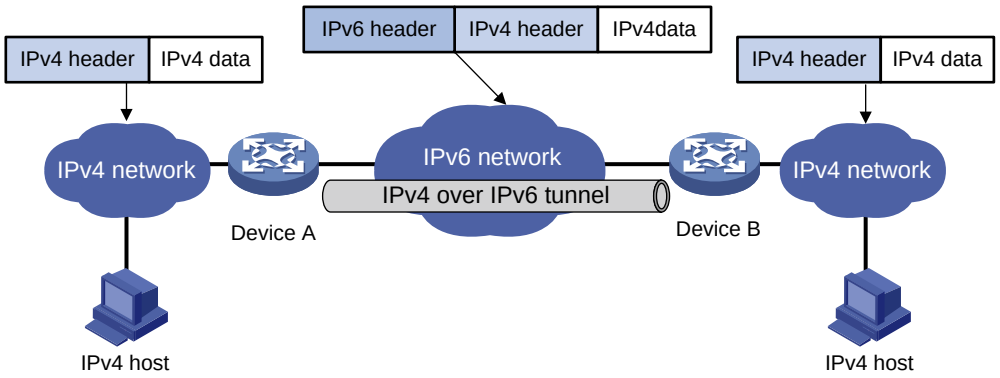
- IPv6 over IPv4 隧道：如图 39 所示，将 IPv6 报文封装到 IPv4 报文中，实现 IPv6 节点跨越 IPv4 网络进行互通。IPv6 over IPv4 隧道包括 IPv6 over IPv4 手动隧道、IPv4 兼容 IPv6 自动隧道、6to4 隧道、ISATAP 隧道和 6RD 隧道。

图39 IPv6 over IPv4 隧道示意图



- IPv4 over IPv6 隧道：如图 40 所示，将 IPv4 报文封装到 IPv6 报文中，实现 IPv4 节点跨越 IPv6 网络进行互通。

图40 IPv4 over IPv6 隧道示意图



隧道技术可以实现不同数据中心网络、不同站点网络穿越骨干网互通，互通过程对于骨干网来说是透明的，多用于广域网。

隧道技术的优点为：

- 原有网络拓扑和路由几乎无需调整，可以短期内快速实现少量 IPv6 站点间的业务互访。
- 仅需对 IPv4 和 IPv6 网络的边界设备进行升级，改造范围小，成本低，且技术比较成熟。

隧道技术的缺点为：

- 无法实现跨协议的应用互访，需与协议转换技术配合使用。
- 隧道技术为软件功能实现，大规模组网时需消耗设备 CPU、内存等资源，设备性能成为瓶颈。
- 运维工作量大，界面不清晰，不推荐大规模部署。

4.3 AFT

4.3.1 AFT 简介

AFT（Address Family Translation，地址族转换）提供了 IPv4 和 IPv6 地址之间的相互转换功能，使 IPv4 网络和 IPv6 网络可以直接通信。

AFT 作用于 IPv4 和 IPv6 网络边缘设备上，所有的地址转换过程都在该设备上实现，对 IPv4 和 IPv6 网络内的用户来说是透明的，即用户不必改变目前网络中主机的配置就可实现 IPv6 网络与 IPv4 网络的通信。

节点不具备升级 IPv6 能力时，在 IPv4 与 IPv6 网络边界点增加协议转换设备，可以快速实现 IPv4/IPv6 节点的跨协议互访。

4.3.2 AFT 前缀转换方式

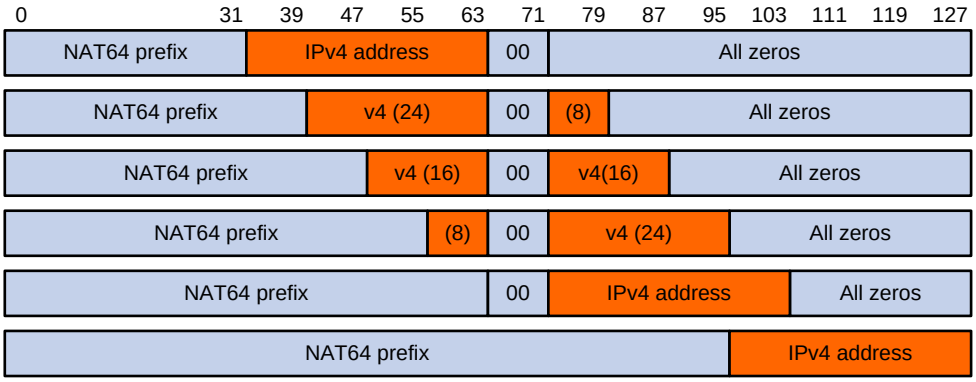
AFT 支持三种前缀转换方式，包括 NAT64 前缀转换、IVI 前缀转换和 General 前缀转换。

1. NAT64 前缀转换

NAT64 前缀是长度为 32、40、48、56、64 或 96 位的 IPv6 地址前缀，用来构造 IPv4 节点在 IPv6 网络中的地址，以便 IPv4 主机与 IPv6 主机通信。网络中并不存在带有 NAT64 前缀的 IPv6 地址的主机。

如图 41 所示，NAT64 前缀长度不同时，地址转换方法有所不同。其中，NAT64 前缀长度为 32、64 和 96 位时，IPv4 地址作为一个整体添加到 IPv6 地址中；NAT64 前缀长度为 40、48 和 56 位时，IPv4 地址被拆分成两部分，分别添加到 64～71 位的前后。64～71 位为保留位，必须设置为 0。

图41 对应 IPv4 地址带有 NAT64 前缀的 IPv6 地址格式



AFT 构造 IPv4 节点在 IPv6 网络中的地址示例如表 9 所示。

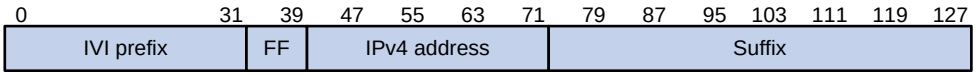
表9 IPv4 地址带有 NAT64 前缀的 IPv6 地址示例

IPv6 前缀	IPv4 地址	嵌入 IPv4 地址的 IPv6 地址
2001:db8::/32	192.0.2.33	2001:db8:c000:221::
2001:db8:100::/40	192.0.2.33	2001:db8:1c0:2:21::
2001:db8:122::/48	192.0.2.33	2001:db8:122:c000:2:2100::
2001:db8:122:300::/56	192.0.2.33	2001:db8:122:3c0:0:221::
2001:db8:122:344::/64	192.0.2.33	2001:db8:122:344:c0:2:2100::
2001:db8:122:344::/96	192.0.2.33	2001:db8:122:344::192.0.2.33

2. IVI 前缀转换

IVI 前缀是长度为 32 位的 IPv6 地址前缀。IVI 地址是 IPv6 主机实际使用的 IPv6 地址，这个 IPv6 地址中内嵌了一个 IPv4 地址，可以用于与 IPv4 主机通信。由 IVI 前缀构成的 IVI 地址格式如图 42 所示。

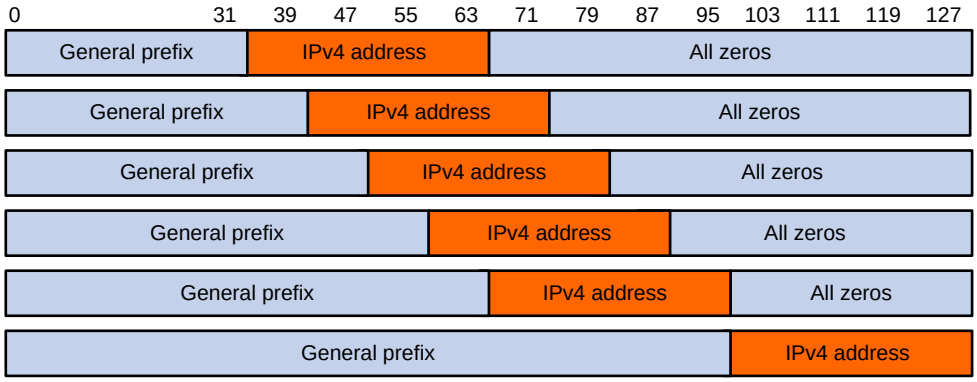
图42 IVI 地址格式



3. General 前缀

General 前缀与 NAT64 前缀类似，都是长度为 32、40、48、56、64 或 96 位的 IPv6 地址前缀，用来构造 IPv4 节点在 IPv6 网络中的地址。如图 43 所示，General 前缀与 NAT64 前缀的区别在于，General 前缀没有 64 到 71 位的 8 位保留位，IPv4 地址作为一个整体添加到 IPv6 地址中。

图43 对应 IPv4 地址带有 General 前缀的 IPv6 地址格式



4.3.3 AFT 的优缺点

AFT 的优点为：

- 仅需要升级 IPv4 和 IPv6 网络边缘设备，改造范围小，短期内可以快速完成 IPv4 向 IPv6 升级改造。
- 业务系统无需做 IPv6 改造升级。

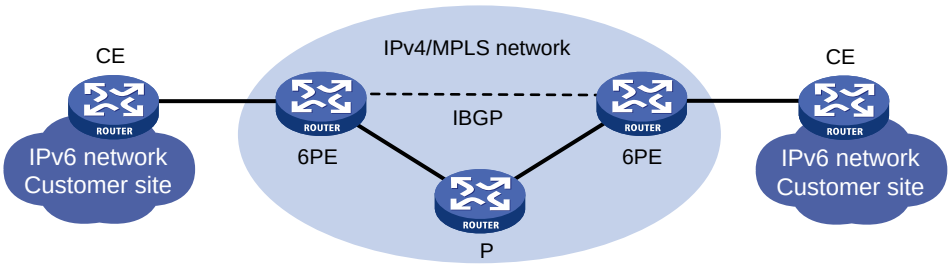
AFT 的缺点为：

- AFT 与业务强耦合，对于部分协议，不仅要对报文头中的源和目的地址进行转换，还需要与 ALG（Application Level Gateway，应用层网关）协同工作，识别出应用层数据载荷中的地址信息并进行地址转换，增加了处理的复杂度。并且，协议转换时，可能造成协议信息的丢失。
- 破坏了 Internet 节点的对等性。
- 溯源困难。

4.4 6PE

如图 44 所示，6PE（IPv6 Provider Edge，IPv6 供应商边缘）是一种过渡技术，它采用 MPLS（Multiprotocol Label Switching，多协议标签交换）技术实现通过 IPv4 骨干网连接隔离的 IPv6 用户网络。当 ISP 希望在自己原有的 IPv4/MPLS 骨干网的基础上，为用户网络提供 IPv6 流量转发能力时，可以采用 6PE 技术方便地达到该目的。

图44 6PE 组网图



6PE 的主要思想是：

- 6PE 设备从 CE（Customer Edge，用户网络边缘）设备接收到用户网络的 IPv6 路由信息后，为该路由信息分配标签，通过 MP-BGP 会话将带有标签的 IPv6 路由信息发布给对端的 6PE 设备。对端 6PE 设备将接收到的 IPv6 路由信息扩散到本地连接的用户网络。从而，实现 IPv6 用户网络之间的路由信息发布。
- 为了隐藏 IPv6 报文、使得 IPv4 骨干网中的设备能够转发 IPv6 用户网络的报文，在 IPv4 骨干网络中需要建立公网隧道。公网隧道可以是 GRE 隧道、MPLS LSP、MPLS TE 隧道等。
- 6PE 设备转发 IPv6 报文时，先为 IPv6 报文封装 IPv6 路由信息对应的标签（内层标签），再为其封装公网隧道对应的标签（外层标签）。骨干网中的设备根据外层标签转发报文，意识不到该报文为 IPv6 报文。对端 6PE 设备接收到报文后，删除内层和外层标签，将原始的 IPv6 报文转发到本地连接的用户网络。

借助 6PE 技术，IPv4 网络运营商仅需对 IPv4 和 IPv6 网络的边界设备进行升级，使其支持 IPv4/IPv6 双协议栈，就可利用自己原有的 IPv4/MPLS 网络为分散的 IPv6 孤岛用户提供接入能力。6PE 技术的优点是部署所需的改造范围小，但缺点是配置复杂，不利于后续进行维护。

4.5 6vPE

6vPE 即 IPv6 MPLS L3VPN，其典型组网环境如图 45 所示。6vPE 组网中，服务提供商骨干网为 IPv4 网络。VPN 内部及 CE 和 PE 之间运行 IPv6 协议，骨干网中 PE 和 P 设备之间运行 IPv4 协议。PE 需要同时支持 IPv4 和 IPv6 协议，连接 CE 的接口上使用 IPv6 协议，连接骨干网的接口上使用 IPv4 协议。PE 从 CE 接收到 IPv6 路由后，为其分配私网标签，并通过 VPNv6 路由将私网标签和 IPv6 路由信息发布给远端 PE。PE 通过 IPv4 骨干网转发 IPv6 报文时，为 IPv6 报文封装私网标签，以实现在 IPv4 网络上透明传输 IPv6 报文，达到 IPv6 网络通过 IPv4 网络互通的目的。

图45 IPv6 MPLS L3VPN 应用组网图

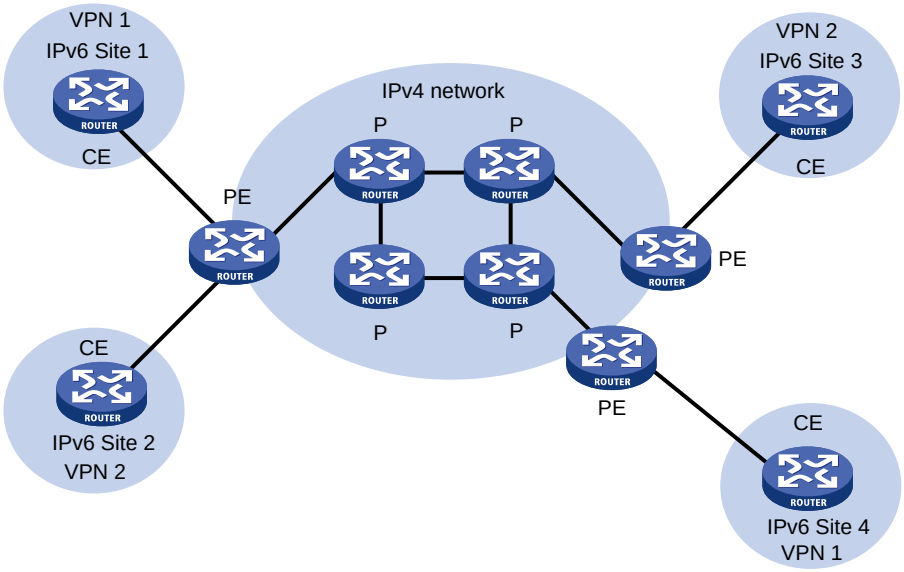
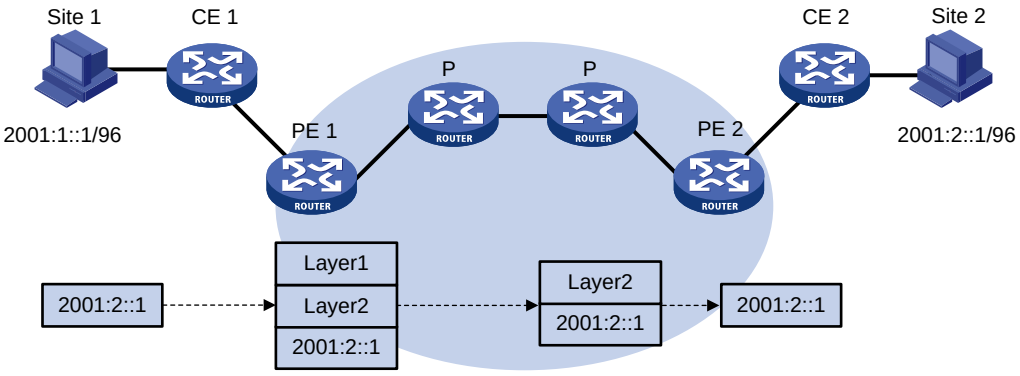


图46 IPv6 MPLS L3VPN 报文转发示意图



如图 46 所示，IPv6 MPLS L3VPN 的报文转发过程为：

(1) Site 1 发出一个目的地址为 2001:2::1 的 IPv6 报文，由 CE 1 将报文发送至 PE 1。

(2) PE 1 根据报文到达的接口及目的地址查找 VPN 实例的路由表项，匹配后将报文转发出去，同时打上公网和私网两层标签。

(3) MPLS 网络利用报文的外层标签，将报文传送到 PE 2。（报文在到达 PE 2 前一跳时已经被剥离外层标签，到达 PE 2 时仅含内层标签）

(4) PE 2 根据内层标签和目的地址查找 VPN 实例的路由表项，确定报文的出接口，将报文转发至 CE 2。

(5) CE 2 根据正常的 IPv6 转发过程将报文传送到目的地。

6vPE 用来通过 IPv4 网络连接孤立的 IPv6 站点，要求骨干网络支持 MPLS L3VPN，部署负责，不利于后期维护。

5 IPv6 演进——IPv6+

5.1 IPv6+概述

IPv6+是面向 5G 和云时代的智能 IP 技术。IPv6+对 IPv6 协议进行了创新，在 IPv6 协议基础上增加了智能识别与控制，具有可编程路径、快速业务发放、自动化运维、质量可视化、SLA 保障和应用感知等特点。

IPv6+技术创新体系的发展分为三个阶段：

- IPv6+1.0：主要为 SRv6 基础特性，包括 TE、VPN、FRR 等目前广泛应用的特性。
- IPv6+2.0：针对 5G 与云时代的新业务与新功能需求，进一步进行一系列的技术创新，包含但不局限于 iFIT、BIER、网络切片、G-SRv6、SFC（service function chaining，业务链）、DetNet 确定性网络等。
- IPv6+3.0：重点是 APN6（application-aware IPv6 networking，感知应用的 IPv6 网络）。APN6 在“IPv6+”2.0 的基础上进一步实现网络能力与业务需求的无缝结合。利用 SRv6 的路径可编程特点，将应用信息（应用标识、对网络性能的需求等）携带在 SRv6 报文中，使网络感知到应用及其需求，以便为其提供相应 SLA 保障。

目前，我司支持的 IPv6+协议创新包括 SRv6、网络切片、iFIT、BIER：

- SRv6（Segment Routing IPv6，IPv6 段路由）：新一代网络承载技术。SRv6 是 IPv6+的关键技术，它通过 IPv6 扩展头，实现网络路径灵活编排。目前，我司支持 SRv6 基础特性、G-SRv6 和 SFC。
- 网络切片：将一个物理网络划分为多个逻辑网络，实现一网多用、业务隔离。
- iFIT（in-situ Flow Information Telemetry）：一种直接测量网络性能指标的检测技术，通过 gRPC 上报检测结果，以实现网络可视化。
- BIER（Bit Index Explicit Replication，位索引显式复制技术）：一种新型的组播转发技术架构，实现 IPv6 组播流量转发的同时，简化了网络协议，并提供了良好的组播业务扩展性。

5.2 SRv6

SRv6 是基于 IPv6 和源路由（Source Routing）的新一代网络承载技术，它简化了传统的复杂网络协议，实现应用级的 SLA 保障。SRv6 具有强大的网络可编程能力，是实现网络自动化的基石。SRv6 能够将网络分片的数据面进行统一，既具备 IPv6 的灵活性和强大的可编程能力，又可以为智能 IP 网络切片、确定性网络、业务链等应用提供强有力的支撑。

5.2.1 SRv6 基本概念

SRv6 是指在 IPv6 网络中使用 Segment Routing，将 IPv6 地址作为 SID，SRv6 节点根据 SID 对报文进行转发。SRv6 将 SID 列表封装在 IPv6 报文的 SRH（Segment Routing Header，SR 报文头）中，以控制报文转发路径。

5.2.2 SRv6 技术优势

SRv6 技术具有如下优势：

- 简化维护

仅需要在源节点上控制和维护路径信息，网络中其他节点不需要维护路径信息。

- 智能控制

  SRv6 基于 SDN 架构设计，跨越了应用和网络之间的鸿沟，能够更好地实现应用驱动网络。SRv6 中转发路径、转发行为、业务类型均可控。

- 部署简单

  SRv6 基于 IGP 和 BGP 扩展实现，无须使用 MPLS 标签，不需要部署标签分发协议，配置简单。

  在 SRv6 网络中，不需要大规模升级网络设备，就可以部署新业务。在 DC（Data Center，数据中心）和 WAN（广域网）中，只需网络边界设备及特定网络节点支持 SRv6，其他设备支持 IPv6 即可。

- 适应 5G 业务需求

  随着 5G 业务的发展，IPv4 地址已经无法满足运营商的网络需求。可通过在运营商网络中部署 SRv6，使所有设备通过 IPv6 地址转发流量，实现 IPv6 化网络，以满足 5G 业务需求。

- 易于实现 VPN 等新业务

  SRv6 定义了多种类型的 SID，不同 SID 具有不同的作用，指示不同的转发动作。通过不同的 SID 操作，可以实现 VPN 等业务处理。

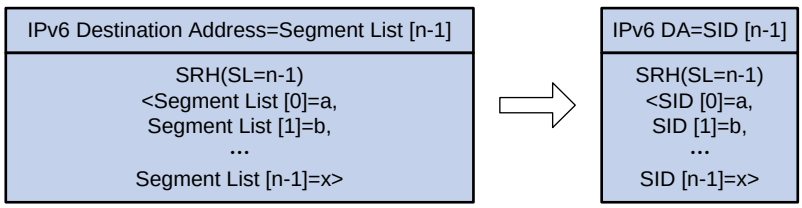
  日后，用户还可以根据实际需要，定义新的 SID 类型，具有很好的扩展性。

5.2.3 SRv6 基本转发机制

如图 47 所示，将 SRv6 报文简化，以便于理解 SRv6 的转发原理，其中：

- IPv6 Destination Address：IPv6 报文的目的地址，简称 IPv6 DA。在普通 IPv6 报文里，IPv6 DA 是固定不变的。在 SRv6 中，IPv6 DA 仅标识当前报文的下一个节点，是不断变换的。
- SRH(SL=n-1)：SRv6 报文的 SID 列表。通过 SL 和 Segment List 字段共同决定 IPv6 DA 的取值。

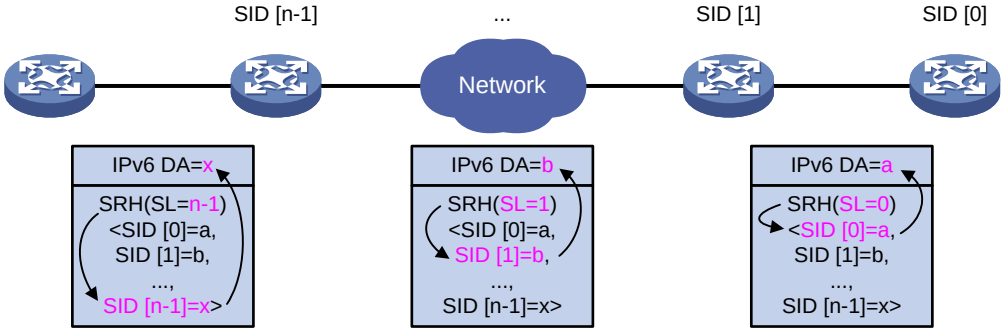
图47 SRv6 报文简化示意图



如图 48 所示，在 SRv6 中，每经过一个 SRv6 节点，SL 字段减 1，IPv6 DA 信息变换一次：

- 如果 SL=n-1，则 IPv6 DA 为 SID [n-1]=x。
- 如果 SL=1，则 IPv6 DA 为 SID [1]=b。
- 如果 SL=0，则 IPv6 DA 为 SID [0]=a。

图48 SRH 处理过程示意图



5.2.4 SRv6 报文转发方式

SRv6 报文支持 SRv6 BE 和 SRv6 TE Policy 两种转发方式：

- SRv6 BE（SRv6 Best Effort）是指通过 IGP 协议发布 Locator 网段，SRv6 网络中的节点按最短路径优先算法计算到达 Locator 网段的最优路由。该路由对应的路径为 SRv6 BE 路径。公网 BGP 路由或者 VPN 实例的 BGP 路由迭代到 SRv6 BE 路径后，可以实现将公网流量或 VPN 流量引入 SRv6 BE 路径。
- SRv6 TE Policy 是指报文的入口节点通过不同的引流方式，将公网流量或 VPN 流量引入 SRv6 TE Policy 转发。SRv6 TE Policy 对应的路径为 SRv6 TE 路径。

5.2.5 G-SRv6

1. 产生背景

在 SRv6 TE Policy 组网场景中，管理员需要将报文转发路径上的 SRv6 节点的 128-bit SRv6 SID 添加到 SRv6 TE Policy 的 SID 列表中。因此，路径越长，SRv6 TE Policy 的 SID 列表中 SRv6 SID 数目越多，SRv6 报文头开销也越大，导致设备转发效率低、芯片处理速度慢。在跨越多个 AS 域的场景中，端到端的 SRv6 SID 数目可能更多，报文开销问题更加严峻。

Generalized SRv6（G-SRv6）通过对 128-bit SRv6 SID 进行压缩，在 SRH 的 Segment List 中封装更短的 SRv6 SID（G-SID），来减少 SRv6 报文头的开销，从而提高 SRv6 报文的转发效率。同时，G-SRv6 支持将 128-bit SRv6 SID 和 G-SID 混合编排到 Segment List 中。

2. G-SRv6 简介

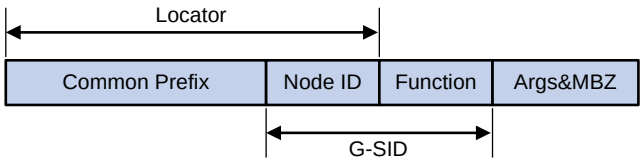
部署 SRv6 时，通常会规划出一个地址块，专门用于 SRv6 SID 的分配，这个地址块称为 SID Space。在一个 SRv6 域中，SRv6 SID 均从 SID Space 中分配，具有相同的前缀（即公共前缀 Common Prefix）。因此，Segment List 中 SRv6 SID 的公共前缀是冗余信息。G-SRv6 将 Segment List 中 SRv6 SID 的 Common Prefix 移除，仅携带 SRv6 SID 中的可变部分，即压缩 SID（G-SID），可以有效减少 SRv6 报文头开销。报文转发过程中，在根据 SRH 头中的 Segment List 替换报文的目的地址时，将 G-SID 与 Common Prefix 拼接形成新的目的地址，继续查表转发。

G-SRv6 对 SRv6 SID 进行压缩时，既要保证高效压缩，又要兼顾网络规模等需求。综合考虑，32 比特是当前较为理想的压缩后 SID 长度。

3. G-SID 格式

如图 49 所示，SRv6 SID 的 Locator 部分可以细分为 Common Prefix 和 Node ID，其中 Common Prefix 表示公共前缀地址；Node ID 表示节点标识。具有相同 Common Prefix 的 SRv6 SID 可以进行压缩，形成 32-bit G-SID。32-bit G-SID 由 128-bit SRv6 SID 中的 Node ID 和 Function 组成。128-bit SRv6 SID 和 32-bit G-SID 的转换关系为：128-bit SRv6 SID ＝ Common Prefix + 32-bit G-SID + 0 (Args&MBZ)

图49 支持压缩的 SRv6 SID 格式图



4. G-SRv6 报文格式

图50 G-SRv6 报文格式示意图

Version	Traffic Class	Flow Label	
Payload Length		Next Header=43	Hop Limit
Source Address(128 bits)			
Destination Address(128 bits)			
Next header	Hdr Ext Len	Routing Type	Segments Left
Last Entry	Flags	Tag	
128-bit SRv6 SID			
G-SID c0	G-SID c1 (COC)	G-SID c2 (COC)	G-SID c3 (COC)
G-SID b0 (COC)	G-SID b1 (COC)	G-SID b2 (COC)	G-SID b3 (COC)
128-bit SRv6 SID			
G-SID a0	G-SID a1 (COC)	G-SID a2 (COC)	G-SID a3 (COC)
128-bit SRv6 SID (COC)			
128-bit SRv6 SID			

说明

如果下一个节点的 SRv6 SID 需要进行压缩，则路由协议在发布本节点的 SRv6 SID 时，会为该 SRv6 SID 添加 COC 标记，标识本 SRv6 SID 之后是 G-SID。报文中不会携带 COC 标记，COC 是 SRv6 SID 本身的转发行为，为了方便理解，在报文结构中标识出 SRv6 SID 是否具有 COC 标记。

如图 50 所示，G-SRv6 可以将 G-SID 和 128-bit SRv6 SID 混合编排在 SRH 的 Segment List 中。为准确定位 G-SID，需要在原本封装 128-bit SRv6 SID 的位置封装 4 个 32-bit G-SID。如果封装的 G-SID 不足 4 个，即不足 128 比特，则需要用 0 补齐，对齐 128 比特。128 比特中封装的 G-SID

称为一组 G-SID。多个连续的 G-SID 组成一段压缩路径，称为 G-SID List。G-SID List 中可以包含一组或多组 G-SID。

G-SID 在 Segment List 中的排列规则为：

(2) G-SID List 的前一个 SRv6 SID 为携带 COC 标记的 128-bit SRv6 SID，标识下一个 SID 为 32-bit G-SID。

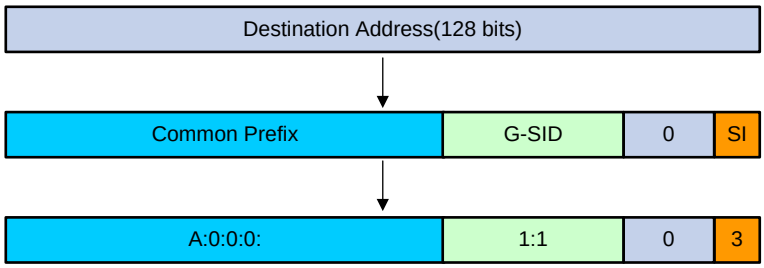
(3) 除 G-SID List 中的最后一个 G-SID 外，其余 G-SID 必须携带 COC 标记，标识下一个 SID 为 32-bit G-SID。

(4) G-SID List 的最后一个 G-SID 必须是未携带 COC 标记的 32-bit G-SID，标识下一个 SID 为 128-bit SRv6 SID。

(5) G-SID List 结束的下一个 SRv6 SID 为 128-bit SRv6 SID，其可以是未携带 COC 标记的 SRv6 SID，也可以是携带 COC 标记的 SRv6 SID。

5. 通过 G-SID 计算目的地址

图51 G-SID 组成示意图



如图 51 所示，使用 G-SID 计算目的地址的方法为将 Segment List 中的 G-SID 与 Common Prefix 拼接形成新的目的地址。其中：

- Common Prefix：公共前缀，由管理员手工配置。
- G-SID：按照 32 比特进行压缩的 SID，从 SRH 中获取。
- SI（SID Index）：用于在一组 G-SID 中定位 G-SID。SI 为目的地址的最低两位，取值为 0～3。每经过一个对 SID 进行压缩的节点，SI 值减 1。如果 SI 值为 0，则将 SL 值减 1。在 Segment List 的一组 G-SID 中，G-SID 按照 SI 从小到大的顺序从左到右依次排列，即最左侧的 G-SID 的 SI 为 0，最右侧的 G-SID 的 SI 为 3。
- 0：若 Common Prefix、G-SID 和 SI 的位数之和不足 128 比特，则中间位使用 0 补齐。

如果 SRv6 节点上管理员部署的 Common Prefix 为 A:0:0:0::/64、SRv6 报文中的当前的 G-SID 为 1:1，该 G-SID 对应的 SI 为 3，则组合成的目的地址为 A:0:0:0:1:1::3。

SRv6 节点收到 G-SRv6 报文后，不同情况下，报文目的地址计算方法为：

- 如果当前报文的目的地址在 Segment List 中为携带 COC 标记的 128-bit SRv6 SID，则表示下一个 SID 为 G-SID，将 SL－1，根据[SL－1]值定位所处的 G-SID 组，并按照上述方法根据[SI＝3]对应的 32-bit G-SID 计算目的地址。
- 如果当前报文的目的地址在 Segment List 中为携带 COC 标记的 32-bit G-SID，则表示下一个 SID 为 G-SID：
  - 如果 SI＞0，则将 SI－1，根据报文当前的 SL 值定位所处的 G-SID 组，并按照上述方法根据[SI－1]对应的 32-bit G-SID 计算目的地址。

- 如果 SI＝0，则将 SL－1、将 SI 值重置，即将 SI 设置为 3，根据报文当前的 SL 值定位所处的 G-SID 组，并按照上述方法根据[SI＝3]对应的 32-bit G-SID 计算目的地址。

- 如果当前报文的目的地址在 Segment List 中是未携带 COC 标记的 32-bit G-SID，则将 SL－1，同时查找[SL－1]对应的 128-bit SRv6 SID，并使用该 SRv6 SID 替换 IPv6 头中的目的地址。
- 如果当前报文的目的地址在 Segment List 中是未携带 COC 标记的 128-bit SRv6 SID，则将 SL－1，同时查找[SL－1]对应的 128-bit SRv6 SID，并使用该 SRv6 SID 替换 IPv6 头中的目的地址。

5.2.6 SRv6 高可靠性

为了保证 SD-WAN 网络中业务流量的稳定，SRv6 提供了高可靠性措施，避免业务流量长时间中断，提高网络质量。

SRv6 提供以下功能保证网络的可靠性：

- TI-LFA FRR（Topology-Independent Loop-free Alternate，拓扑无关无环备份快速重路由）：高保护率的 FRR 保护能力，TI-LFA FRR 原理上支持任意拓扑保护，能够弥补传统隧道保护技术的不足。
- SRv6 防微环：解决全互联组网中 IGP 协议在无序收敛时产生的环路，支持正切防微环和回切防微环，消除微环导致的网络丢包、时延抖动和报文乱序等一系列问题。
- 中间节点保护：解决 SRv6 TE Policy 场景由于严格节点约束导致的 TI-LFA FRR 保护失效问题。
- 尾节点保护：在双归接入场景中，解决 SRv6 TE Policy 的尾节点发生单点故障，引起的报文转发失败问题。

5.2.7 SRv6 VPN

传统 VPN 网络中通过部署 LDP/RSVP-TE 等标签分发协议，在公网中建立虚拟专用通信网络。这种方式部署复杂，维护成本较高。通过在公网部署 SRv6 VPN 可以解决上述问题。SRv6 VPN 是通过 SRv6 隧道承载 IPv6 网络中的 VPN 业务的技术，控制平面采用 MP-BGP 通告 VPN 路由信息，数据平面采用 SRv6 封装方式转发报文。租户的物理站点分散在不同位置时，SRv6 VPN 可以基于已有的服务提供商或企业 IP 网络，为同一租户的不同物理站点提供二层或三层互联。

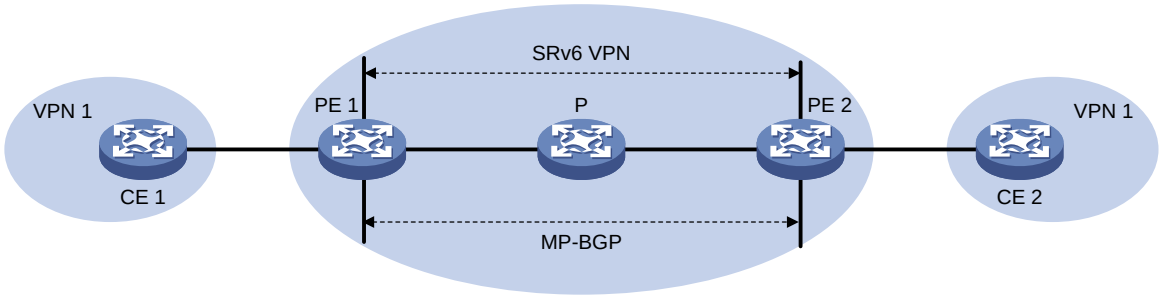
根据 VPN 业务种类，SRv6 VPN 分为：

- L3VPN 业务：IP L3VPN over SRv6 和 EVPN L3VPN over SRv6
- L2VPN 业务：EVPN VPWS over SRv6 和 EVPN VPLS over SRv6

1. IP L3VPN over SRv6

如图 52 所示，IP L3VPN over SRv6 通过 MP-BGP 在 IPv6 骨干网上发布用户站点的 IPv4/IPv6 私网路由，使用 PE 间的 SRv6 路径承载私网报文，从而实现通过 IPv6 骨干网连接属于同一个 VPN、位于不同地理位置的用户。

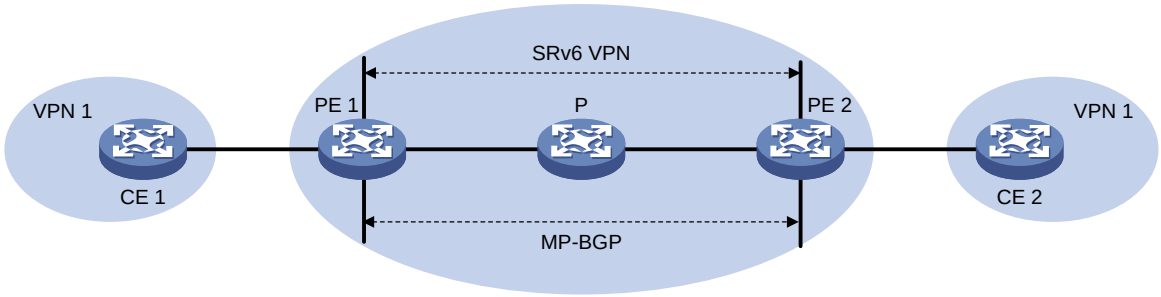
图52 IP L3VPN over SRv6 组网示意图



2. EVPN L3VPN over SRv6

如图 53 所示，EVPN L3VPN over SRv6 通过 MP-BGP 在 IPv6 骨干网上使用 EVPN 的 IP 前缀路由发布用户站点的 IPv4/IPv6 私网路由，使用 PE 间的 SRv6 路径承载私网报文，从而实现通过 IPv6 骨干网连接属于同一个 VPN、位于不同地理位置的用户。

图53 EVPN L3VPN over SRv6 组网示意图

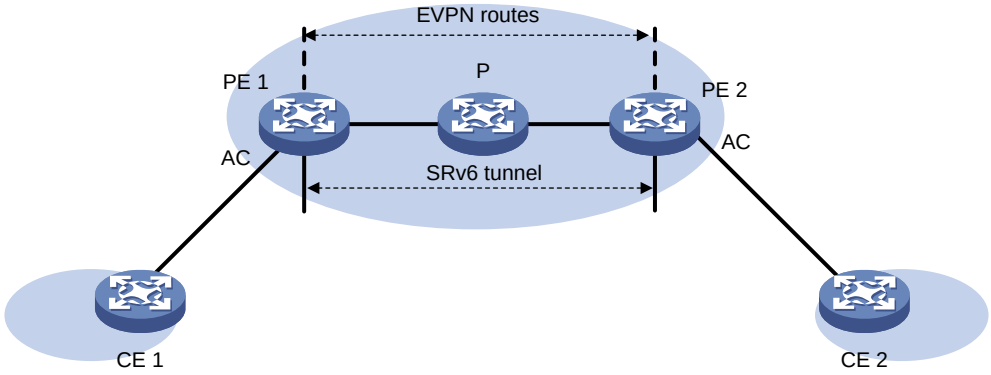


3. EVPN VPWS over SRv6

EVPN VPWS over SRv6 是指通过 SRv6 隧道承载 EVPN VPWS 业务，通过 IPv6 网络透明传输用户二层数据，实现用户网络穿越 IPv6 网络建立点到点连接。

如图 54 所示，PE 之间通过 EVPN 路由发布 SRv6 SID，建立 SRv6 隧道。该 SRv6 隧道作为 PW，封装并转发站点网络之间的二层数据报文。

图54 EVPN VPWS over SRv6 组网示意图

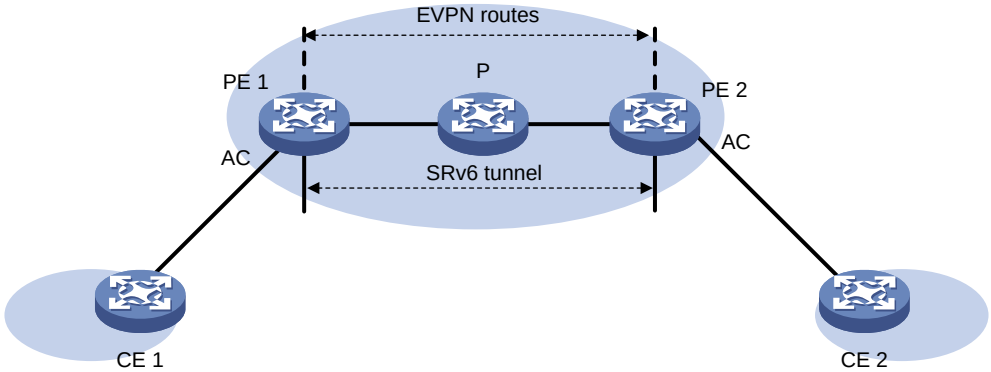


4. EVPN VPLS over SRv6

EVPN VPLS over SRv6 是指通过 SRv6 隧道承载 EVPN VPLS 业务，通过 IPv6 网络透明传输用户二层数据，实现用户网络穿越 IPv6 网络建立点到多点连接。

如图 55 所示，PE 之间通过 EVPN 路由发布 SRv6 SID，建立 SRv6 隧道。该 SRv6 隧道作为 PW，封装并转发站点网络之间的二层数据报文。

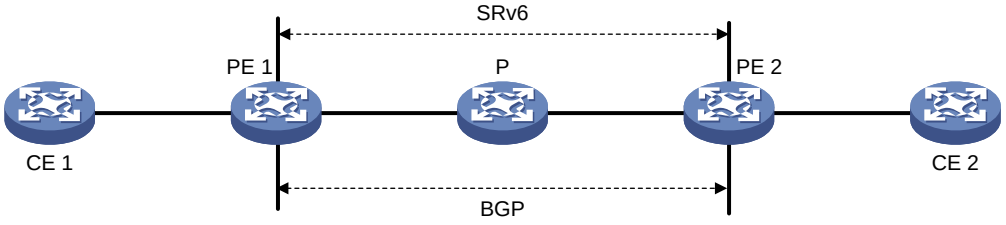
图55 EVPN VPLS over SRv6 组网示意图



5. 公网 IP over SRv6

如图 56 所示，公网 IP over SRv6 通过 MP-BGP 在 IPv6 骨干网上发布用户站点的 IPv4/IPv6 路由，使用 PE 间的 SRv6 路径承载用户报文，从而实现通过 IPv6 骨干网连接位于不同地理位置的用户。

图56 公网 IP over SRv6 组网示意图



5.3 网络切片

5.3.1 网络切片概述

网络切片是指在同一个物理网络的基础上，网络管理员通过各种切片技术为特定业务或用户上划分出多个逻辑网络，即切片网络。每个切片网络都有自己的网络拓扑、SLA（Service Level Agreement）需求、安全和可靠性要求。切片网络最大化地利用现网的网络设施资源，为不同业务、行业或用户提供差异化网络服务。

5.3.2 网络切片的价值

网络切片的价值体现在如下几个方面：

- 满足差异化 SLA 需求：对于运营商或者大型企业而言，当前不同业务通过一张网络承载，不断涌现的新业务对这张网络提出了差异化 SLA 的需求，例如自动驾驶业务对时延，抖动的要求十分严格，但带宽需求不大，而 VR 和高清视频等业务又对网络带宽需求极大，对时延并无特殊需求。传统的物理网络无法满足差异化 SLA 要求，而建设独立的专网成本过高，部署网络切片方案为不同业务按需提供不同切片网络可以满足上述需求。
- 满足网络资源隔离的需求：一些行业用户需要安全可靠且独享的网络资源，运营商也希望为不同等级的用户提供安全隔离措施，避免部分普通客户抢占网络资源，造成其他优质用户体验下降问题，网络切片方案可以在数据平面、控制平面和管理层面为不同用户分配不同资源。
- 满足灵活定制拓扑的需求：随着云网融合技术的发展，虚拟机可以跨数据中心随时迁移，网络的连接关系更加灵活复杂，网络切片方案中通过部署 Flex-Algo 技术满足网络拓扑灵活动态的变化需求。

5.3.3 网络切片的技术方案

网络切片并非特指某一种网络技术，而是利用多种网络技术实现的一整套解决方案。为了实现在物理网络上划分逻辑网络的功能，满足不同用户和业务的资源隔离、差异化 SLA 以及灵活拓扑的需求，网络切片方案中包含但不限于表 10 中所列的技术。

表10 网络切片技术

技术名称	实现层级	说明
MDC	• 管理平面 • 控制平面 • 数据平面	通过虚拟化技术将一台物理设备划分成多台逻辑设备，每台逻辑设备就称为一台MDC（Multitenant Device Context，多租户设备环境）。 每台MDC拥有自己专属的软硬件资源，独立运行，独立转发，独立提供业务。对于用户来说，每台MDC就是一台独立的物理设备。MDC之间相互隔离，不能直接通信，具有很高的安全性。
Flex-Algo	控制平面	Flex-Algo（Flexible Algorithm，灵活算法）是一种在IGP协议基础上运行的灵活算法，它允许用户自定义IGP路径算法的度量类型，例如Cost开销、链路时延值或TE度量值，利用SPF算法计算到达目的地址的最短路径。 计算最短路径时，Flex-Algo还允许用户使用的链路的亲和属性和SRLG（Shared Risk Link Group，共享风险链路组）作为约束条件来限制最终拓扑必须包含或排除某些链路。

技术名称	实现层级	说明
		因此，参与不同Flex-Algo算法的网元可以组成多个独立的逻辑拓扑，物理网络中部署多个Flex-Algo算法可以按需划分成多个独立的网络切片。
FlexE	数据平面	FlexE（Flexible Ethernet，灵活的以太网）技术基于高速以太网接口，通过以太网MAC速率和PHY速率的解耦，实现灵活控制接口速率。 FlexE通过一个或捆绑多个IEEE 802.3标准的高速物理接口提供大带宽，再根据业务带宽需求，将上述物理接口的总带宽灵活分配给各FlexE业务接口。每一个FlexE业务接口就可以为切片网络转发数据。
子接口切片	数据平面	子接口切片是一种小粒度的网络切片技术。通过在高速率端口上创建子接口，并为这些子接口配置子接口切片带宽，利用接口的队列资源，实现子接口上数据转发的隔离。这些配置了子接口切片带宽的子接口称为切片子接口。切片子接口独享为其分配的带宽，并使用独立的队列进行调度。
Slice ID切片	数据平面	基于Slice ID的网络切片是一种应用在SRv6组网场景中的网络切片技术方案，它通过全局唯一的Slice ID来标识和划分切片网络。 在切片网络中转发的业务报文将携带Slice ID，设备转发该报文时先查询FIB表找到出接口，再根据Slice ID从对应出接口上切片通道转发报文。

采用 Slice ID 切片技术的网络切片方案因为支持的切片数量多（可达千级），配置实现简单，转发接口所需 SRv6 SID 少（所有网络切片仅需一套 SRv6 Locator 资源）等优势成为网络切片当前推荐方案。下面仅以基于 Slice ID 的网络切片方案为例，介绍网络切片的基本原理。

5.3.4 基于 Slice ID 的网络切片实现原理

Slice ID 是实现网络切片的关键：

- 本方案通过在设备上配置全局唯一的 Slice ID 来创建切片实例，即虚拟设备，实现物理设备到切片网络中的映射；
- 在设备接口上的配置 Slice ID 来创建不同切片网络使用的的数据转发通道，并为该数据通道分配独享带宽资源；
- 最后，由同一 Slice ID 来标识的虚拟设备和数据转发通道组成切片网络。

在数据转发过程中，网络切片里转发的报文也必须携带 Slice ID。同时 SRv6 TE Policy 需要与 Slice ID 关联。SRv6 TE Policy 可以通过如下方法与 Slice ID 关联：

- 通过配置手工指定。
- 从对等体发布的 BGP IPv6 SR Policy 路由中学习。

以 L3VPNv4 over SRv6 TE Policy 组网为例，介绍基于 Slice ID 的网络切片方案的报文转发过程。

在 SRv6 网络中存在 Slice ID 1 和 Slice ID 2 分别用来表示两个切片网络。PE 1、P 和 PE 2 的转发接口上均存在分别与 Slice ID 1 和 Slice ID 2 关联的网络切片通道 1 和网络切片通道 2。VPN 1 中两个站点 CE 1 和 CE 2 之间的流量使用 SRv6 TE Policy 承载，SRv6 TE Policy 关联了 Slice ID 1。CE 1 访问 CE 2 的报文转发过程为：

(1) CE 1 向 PE 1 发送 IPv4 单播报文。PE 1 接收到 CE 1 发送的报文之后，查找 VPN 实例路由表，该路由的出接口为 SRv6 TE Policy。PE 1 为报文封装如下信息：

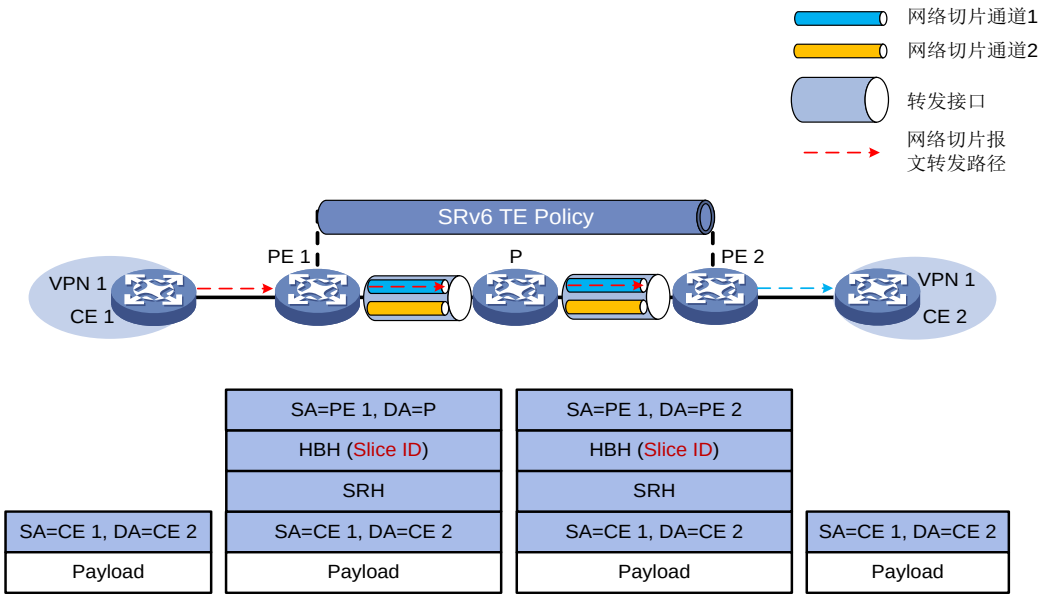
- 封装 SRH 头，在 SRH 头中携带 SRv6 TE Policy 的 SID List。
- 封装 HBH 扩展头，在 HBH 扩展头中携带 SRv6 TE Policy 关联的 Slice ID 1。
- 封装 IPv6 基本报文头。

(2) PE 1 将报文转发给 P，转发时根据 Slice ID 信息在出接口上查找与其对应的网络切片通道，并通过该通道转发报文。

(3) 中间 P 备根据 SRH 信息转发报文，转发时根据 Slice ID 在出接口上查找与其对应的网络切片通道，并通过该通道转发报文。

(4) 报文到达尾节点 PE 2 之后，PE 2 使用报文的 IPv6 目的地址查找 Local SID 表，匹配到 End SID，PE 2 将报文 SL 减 1，IPv6 的目的地址更新为 End.DT4 SID。PE 2 根据 End.DT4 SID 查找 Local SID 表，执行 End.DT4 SID 对应的转发动作，即解封装报文去掉 IPv6 报文头，并根据 End.DT4 SID 匹配 VPN 1，在 VPN 1 的路由表中，查表转发，将报文发送给 CE 2。

图57 基于 Slice ID 的网络切片方案的报文转发过程



5.4 iFIT

5.4.1 iFIT 概述

iFIT 是一种应用于 MPLS（Multiprotocol Label Switching，多协议标签交换）、SR-MPLS（Segment Routing MPLS，MPLS 段路由）、SRv6、G-SRv6（Generalized SRv6，通用 SRv6）和 G-BIER（Generalized BIER，通用位索引显式复制）网络的、测量网络性能指标的测量技术，它直接测量业务报文的真实丢包率和时延等参数，具有部署方便、统计精度高等优点。

说明

iFIT 对 G-SRv6 和 G-BIER 网络的支持情况与设备的型号有关，请以设备的实际情况为准。

5.4.2 技术优点

相较于传统丢包测量技术，iFIT 具有以下优势：

- 检测精度高：直接对业务报文进行测量，测量数据可以真实反映网络质量状况。
- 部署简单：中下游设备可以根据 iFIT 报文生成测量信息。
- 快速定位故障功能：iFIT 提供了随流检测功能，可以真正实时地检测用户流的时延，丢包情况。
- 可视化功能：iFIT 通过可视化界面展示性能数据，具备快速发现故障点的能力。
- 支持路径自发现功能：iFIT 在网络中的入节点对于用户关心的业务流程增加报文头，下游设备可以根据 iFIT 报文头自动识别该业务流并生成统计测量信息；分析器可以通过该功能感知业务流量在网络中的实时路径。
- 基于硬件实现，对于网络影响较小，可扩展性强。

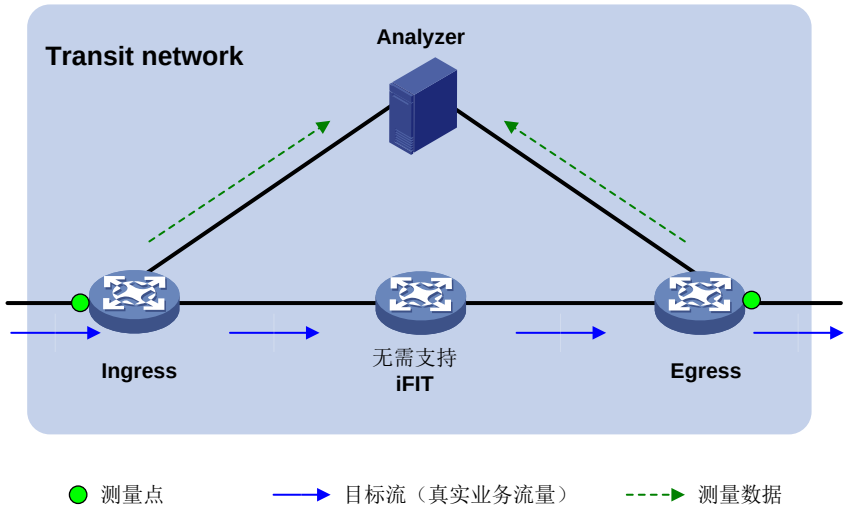
5.4.3 应用场景

iFIT 支持以下两种测量类型：端到端测量和逐点测量，这两种测量类型适用于不同应用场景。

- 端到端测量

  当用户希望测量整个网络的丢包和时延性能时，可以选择端到端测量类型。端到端测量会测量流量在进入网络的设备（流量入口）和离开网络的设备（流量出口）之间是否存在丢包以及时延参数。如图 58 所示，iFIT 可用于直接测量流量从 Ingress（入节点）到达 Egress（出节点）时，是否有丢包、时延，以及丢包率和时延值。

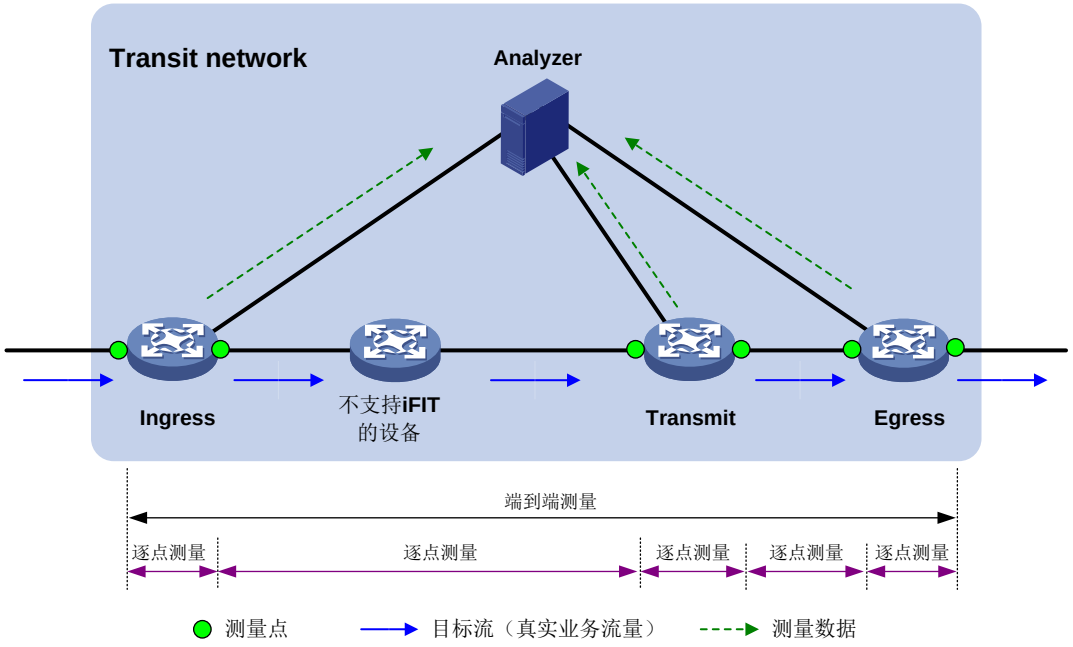
  图58 端到端测量应用示意图



- 逐点测量

当用户希望准确定位每个网络节点的丢包和时延性能时，可以选择逐跳测量类型。当根据测量结果发现端到端统计场景有丢包或者时延不满足业务要求时，可以将端到端之间的网络划分为多个更小的测量区段，测量每两个网元之间是否存在丢包、时延值，进一步定位影响网络性能的网元位置。如图 59 所示，iFIT 可同时测量流量从 Ingress 到达 Egress 时，Ingress 和 Transmit（中间节点）之间、Transmit 和 Egress 之间任意两个接口间是否有丢包、时延，以及丢包率和时延值。

图59 逐点测量应用示意图



5.4.4 网络框架

如 5.4.3 图 58、5.4.3 图 59 所示，iFIT 网络框架中主要涉及三个对象：目标流、目标流穿越的网络（Transit network）和统计系统。

1. 目标流

目标流是 iFIT 统计的目标对象。根据生成方式不同目标流分为静态目标流和动态目标流两种。

- 静态目标流：静态目标流是入节点上手工配置的流匹配规则，在入节点上使用命令行配置完 iFIT 静态目标流，且开启 iFIT 测量功能后，入节点会生成一个 iFIT 静态目标流。设备支持的匹配规则包括五元组（源 IP 地址/网段、源端口、目的 IP 地址/网段、目的端口、协议类型）、DSCP、VPN 和下一跳参数。

  iFIT 报文头中包含 DeviceID、FlowID、测量周期、测量类型、是否需要测量时延、是否需要测量丢包等重要参数。其中：

  - DeviceID：设备的标识。在 iFIT 测量网络中，设备 ID 用来唯一标识一台设备
  - FlowID：FlowID 由入节点自动生成，会封装到 iFIT 报文头中传递给中间节点和出节点，用于在 iFIT 测量网络中与设备标识 DeviceID 一起唯一的标识一条目标流。

- 测量周期：设备按周期进行 iFIT 测量，从开始一次测量，到收集并上报该次测量数据的时间间隔称为一个测量周期。
- 测量类型：表示本次测量是端到端测量还是逐点测量。

- 动态目标流：动态目标流是设备动态学习到的业务流。
  - 对于入节点，如果设备收到的报文匹配静态目标流的配置，则入节点会生成一个和静态流 Flow ID 相同的动态目标流。
  - 对于中间节点和出节点，则通过解析收到的报文，根据报文中携带的 iFIT 报文头动态学习生成动态目标流。

  设备以 iFIT 报文头中的“DeviceID+FlowID”作为划分动态目标流的依据。如果在指定时间内没有收到相同“DeviceID+FlowID”的报文，则认为该动态目标流已经老化，设备会将该动态目标流删除。


说明

iFIT 对 DeviceID 的支持情况与设备的型号有关，请以设备的实际情况为准。

- 对于支持 DeviceID 参数的产品，“DeviceID+FlowID”用于标识一条 iFIT 流。
- 对于不支持 DeviceID 参数的产品，用 FlowID 标识一条 iFIT 流。

2. 测量点

测量点（Detection point）：实施 iFIT 测量的接口。用户可根据测量需求指定测量点。

3. 目标流穿越网络

目标流穿越网络是传输目标流的网络，目标流既不在该网络内产生，也不在该网络内终结。目前支持的目标流穿越网络只能是三层网络。网络内的设备必须路由可达。

4. 统计系统

统计系统指的是完成 iFIT 性能统计的所有设备的集合。它包含了以下角色：

- 入节点（Ingress）：目标流进入目标流穿越网络的设备，它负责对目标流进行筛选，为目标流添加 iFIT 报文头，收集目标流的统计数据并上报给 Analyzer。
- 中间节点（Transmit）：根据报文是否包含 iFIT 报文头来判断是否为 iFIT 目标流，对于 iFIT 目标流，再根据 iFIT 头中携带的测量类型，决定是否需要收集目标流的统计数据并上报给 Analyzer。
- 出节点（Egress）：根据报文是否包含 iFIT 报文头来判断是否为 iFIT 目标流，对于 iFIT 目标流，收集目标流的统计数据并上报给 Analyzer，去掉报文中的 iFIT 报文头。
- 分析器（Analyzer）：负责收集入节点、中间节点、出节点上送的统计数据并完成数据的汇总和计算。

5.4.5 工作机制

1. 时间同步机制

iFIT 以时间同步为基础。在测量开始前，要求所有参与 iFIT 测量的设备时间已经同步，从而确保各个设备能够基于相同的周期进行报文统计和上报。如果时间不同步，会导致 iFIT 计算结果不准确。

分析器和 iFIT 设备的时间同步与否不影响计算结果，但为了便于管理和维护，建议分析器和所有 iFIT 设备的时间均保持同步。iFIT 使用 PTP（Precision Time Protocol，精确时间协议）协议进行时间同步。关于 PTP 功能的具体描述和配置请参见“网络管理和监控配置指导”中的“PTP”。

2. 丢包测量机制

iFIT 丢包计算依据报文守恒原理，即每个周期内、从入节点进入的报文总数应该等于出节点发送的报文总数。如果不相等，则说明目标流穿越网络内存在丢包现象。

3. 时延测量机制

iFIT 时延测量机制原理为：每个测量点会记录目标流中每个报文经过自己的时间戳 t0；在下游测量点匹配该报文，并记录该报文经过自己的时间戳 t1。最终两个测量点分别将两个时间戳上报给分析器，由分析器计算时延。

4. 测量数据上报机制

iFIT 采用 gRPC（Google Remote Procedure Call，Google 远程过程调用）协议将测量数据从 iFIT 设备推送给 iFIT 分析器。

iFIT 目前支持 gRPC Dial-out 模式，iFIT 设备作为 gRPC 客户端，iFIT 分析器作为 gRPC 服务器（在 gRPC 协议中也称为采集器）。

设备支持按照以下两个周期将测量数据从 iFIT 设备推送给 iFIT 分析器，请根据需要选择使用一种即可：

- gRPC 订阅周期：如果管理员配置 gRPC 订阅时配置了采样周期，则不管采样路径是周期采样路径还是事件触发采样路径，设备主动和分析器建立 gRPC 连接后，均会按照 gRPC 订阅周期将设备上订阅的 iFIT 统计数据推送给分析器。
- iFIT 测量周期：如果管理员配置 gRPC 订阅时未配置采样周期，且采样路径是周期采样路径，因为缺少采样周期配置，设备无法将 iFIT 统计数据推送给分析器；如果管理员配置 gRPC 订阅时未配置采样周期，且采样路径是事件触发采样路径，设备主动和分析器建立 gRPC 连接后，设备按照 iFIT 测量周期定时将设备上订阅的 iFIT 统计数据推送给分析器。

说明

- iFIT 的周期采样路径为 ifit/flowstatistics/flowstatistic。
- iFIT 的事件触发采样路径为 insuitoam/measurereport（该路径包含了 iFIT 的测量结果）和 insuitoam/flowinfo（该路径包含了 iFIT 流的信息）。

关于 gRPC 的详细介绍请参见“Telemetry 配置参考”中的“gRPC”。关于 iFIT 采样路径的详细介绍请参见 iFIT 的 NETCONF API 文档。

5. 工作流程

下面以逐点测量场景为例，说明 iFIT 的工作机制。（端到端测量的流程与此类似，只是不需要部署中间节点。）

以图 60 所示组网为例，目标流穿越网络中有四台设备，其中三台支持 iFIT，在这三台设备上部署 iFIT 功能，iFIT 的工作流程如下：

(1) Analyzer 和所有 iFIT 设备之间通过 PTP 协议完成时间的同步。

(2) iFIT 设备对目标流报文进行 iFIT 处理。

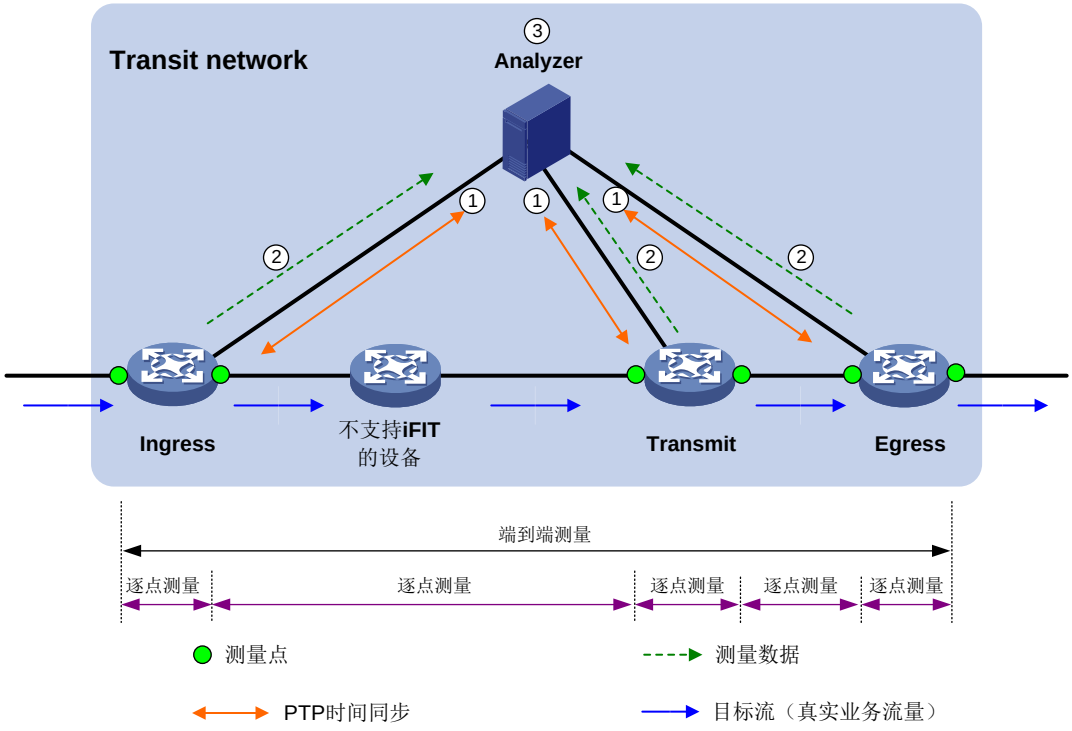
a. 在目标流穿越网络的入接口（入节点上用户手工绑定目标流的接口），iFIT 会解析流经该接口的报文，按照规则完成目标流的匹配，给目标流报文添加 iFIT 报文头，统计目标流报文个数，同时按周期将报文计数和时间戳等信息通过 gRPC 连接上报给分析器。

b. 在目标流穿越网络的传输接口（在目标流穿越网络中支持 iFIT 的设备上，目标流的流入接口和流出接口），对于包含 iFIT 报文头的报文，iFIT 会统计这些报文的个数，同时按周期将报文计数和时间戳等信息通过 gRPC 连接上报给分析器。

c. 在目标流穿越网络的出接口（目标流离开目标流穿越网络的接口，可能为出节点上目标流的流入接口，也可能是目标流的流出接口，具体是流入接口还是流出接口不同设备支持情况不同，请以设备实际情况为准），iFIT 会解析流经该接口的报文，按照规则完成目标流的匹配，对于包含 iFIT 报文头的报文，iFIT 统计目标流报文个数，同时按周期将报文计数和时间戳等信息通过 gRPC 连接上报给分析器，去掉目标流报文中的 iFIT 报文头，继续转发。

(3) 分析器对相同周期、相同实例、相同流量进行丢包分析，计算时延。

图60 iFIT 工作机制示意图



5.5 BIER

5.5.1 概述

1. 产生背景

传统 IP 组播和组播 VPN 技术中，设备需要为每条组播流量分别建立组播分发树，分发树中的每一个节点都需要感知组播业务，并保留组播流状态。例如，公网 PIM 组播中，需要为每条组播流量建立一个 PIM 的组播分发树；在 NG MVPN 中，需要为每条组播流量建立 P2MP 隧道，也相当于建

立一个 P2MP 组播树。随着组播业务的大规模部署，待维护的组播分发树的数量也急剧增加，组播节点上需要保留大量的组播流状态，当网络发生变化的时候，会导致组播表项收敛缓慢。

同时，单播路由协议、组播路由协议、MPLS 协议等多协议在承载网络上并存，增加了承载网络控制平面的复杂度，使得故障收敛速度慢，运维困难，难以向 SDN 架构网络演进。

BIER 是一种新型的组播转发技术架构，通过将组播报文要到达的目的节点集合以 BS（Bit String，位串）的方式封装在报文头部发送，使得网络中间节点无需感知组播业务和维护组播流状态，可以较好地解决传统 IP 组播技术存在的问题，提供了良好的组播业务扩展性。

在 BIER 网络中，组播报文的转发依靠 BFR（Bit Forwarding Router，位转发路由器）上通过 BIER 技术建立的 BIFT（Bit Index Forwarding Table，位索引转发表），实现组播报文只需根据位串进行复制和转发。

目前，我司支持 G-BIER（Generalized BIER，通用位索引显式复制）和 BIERv6（Bit Index Explicit Replication IPv6 Encapsulation，IPv6 封装的比特索引显式复制）两种封装类型。

2. 技术优点

BIER 具有如下几方面的技术优点：

- 良好的组播业务扩展性

  BFR 上采用 BIER 技术建立的 BIFT 是独立于具体的组播业务的公共转发表，使得网络中间节点无需感知组播业务，不需要维护特定组播业务的组播流状态。公网组播和私网组播报文均可通过 BIFT 转发，具有良好的组播业务扩展性。

- 简化业务部署和运维

  由于网络中间节点不感知组播业务，因此部署组播业务不涉及中间节点，组播业务变化对中间节点没有影响，简化了网络的部署和运维。

- 简化承载网络的控制平面

  在承载网络的中间节点上，不需要运行 PIM 协议，控制平面协议统一为单播路由协议 IGP 和 BGP，简化了承载网络的控制平面协议。

- 利于 SDN 架构网络演进

  部署组播业务不需要操作网络中间节点，只需在入口节点为组播报文添加上指示后续组播复制的 BIER 封装。BIER 封装中携带标识组播出口节点的位串，中间节点根据位串实现组播复制和转发，有利于 SDN 架构网络的演进。

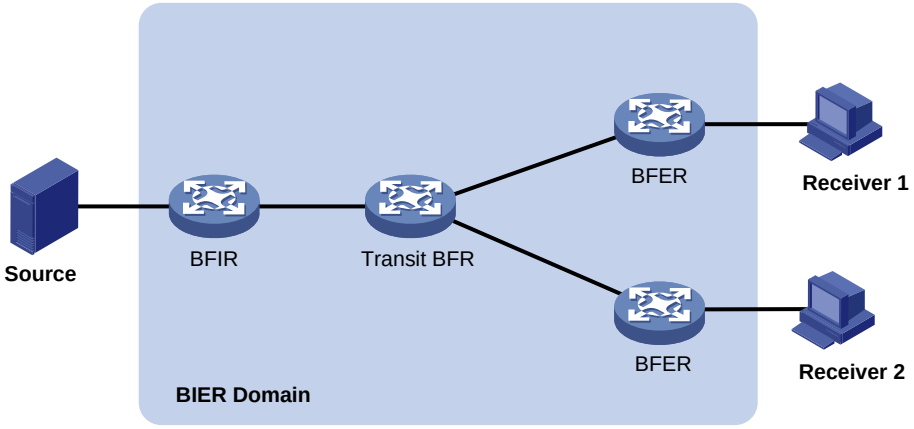
5.5.2 网络模型

BIER 网络的基本元素为支持 BIER 转发的 BFR（Bit Forwarding Router，位转发路由器）。如图 61 所示，BIER 的典型网络模型中包括以下几个部分：

- BFIR（Bit Forwarding Ingress Router，位转发入口路由器）：组播数据流量进入 BIER 域的节点，负责对进入 BIER 网络的报文进行 BIER 封装。
- Transit BFR：组播数据流量在 BIER 域中转发的中间节点，负责对 BIER 报文进行转发。
- BFER（Bit Forwarding Egress Router，位转发出口路由器）：组播数据流量出 BIER 域的节点，负责对 BIER 报文进行解封装，并转发给组播接收者。

其中，BFIR 和 BFER 统称为 BIER 边缘设备。

图61 BIER 典型网络模型



5.5.3 基本概念

- BIER 域和 BIER 子域：在一个路由域或者管理域内所有 BFR 的集合称为 BIER 域（Domain）。一个 BIER 域可以划分为一个或者多个 BIER 子域（Sub-domain），Sub-domain 也可简称为 SD。每个 BIER 子域通过一个唯一的 Sub-domain ID 来标识。
- BFR ID：用来在一个 BIER 子域中唯一标识一台 BIER 边缘设备，Transit BFR 无需配置 BFR ID。
- BFR prefix：相当于路由协议中的 Router ID，用来标识 BFR。在同一个 BIER 子域中，每个 BFR 必须配置唯一的 BFR 前缀，且该前缀必须是 BIER 子域内路由可达的。目前 BFR prefix 只支持配置为 Loopback 口的地址。
- BS（Bit String，比特串）：BIER 封装用一个特定长度的 BS 来表示 BIER 报文的目的边缘设备。Bit String 从最右边开始，每一个比特位对应一个 BFR ID。比特位置 1，表示该比特位对应的 BFR ID 所标识的 BIER 边缘设备，为组播报文转发的目的边缘设备。
- BSL（Bit String Length，比特串长度）：用来表示 BIER 封装中的比特串长度。
- SI（Set Identifier，集标识）：当一个 BIER 子域内使用的 BSL 长度不足以表示该子域内配置的 BFR ID 的最大值时，需要将 Bit String 分成不同的集合，每个集合通过 SI 来标识。比如，BIER 子域内 BFR ID 最大值为 1024，假如 BSL 设置为 256，就需要将 BIER 子域分为四个集合，分别为 SI 0，SI 1、SI 2 和 SI 3。
- BIRT（Bit Index Routing Table，位索引路由表）：BFR 上结合 IGP/BGP 协议交互的 BIER 属性信息（Sub-domain ID、BSL、BFR ID）与 IGP/BGP 路由信息生成的 BIER 路由表项，用于指导 BIER 报文的转发。
- F-BM（Forwarding-Bit Mask，转发位串掩码）：用来表示 BFR 往下一跳邻居复制发送组播报文时，通过该邻居能到达的 BIER 子域边缘节点集合。F-BM 是 BFR 通过将该邻居所能到达的所有 BIER 子域边缘节点的 Bit String 进行“或”操作后得到。
- BIFT（Bit Index Forwarding Table，位索引转发表）：BIER 子域内的组播流量通过查询 BIFT 来实现逐跳转发。每张 BIFT 都由三元组（BSL，SD，SI）确定。BIFT 是 BFR 将 BIRT 表项中经过相同邻居不同表项进行合并生成，每条表项记录了一个下一跳邻居和对应的 F-BM。

5.5.4 三层网络架构

IETF RFC 8279 将 BIER 网络架构分为 Underlay、BIER 和 Overlay 三层。

(1) Underlay 层

Underlay 层为传统 IP 路由层，通过 IGP 协议（目前仅支持 IS-IS）的扩展 TLV 属性携带 BFR 的 BIER 属性信息在 BIER 子域内进行泛洪。BFR 根据 IGP 算法生成到本子域内其它 BFR 前缀的路由，也就是到每个 BFR 的路由，从而建立 BIER 子域内节点之间的邻居关系以及节点之间的最佳转发路径。

(2) BIER 层

BIER 层作为 BIER 转发的核心层，在控制平面对 BIER 转发所需的 IGP 协议进行了扩展，用于生成指导组播报文在 BIER 域内完成转发的 BIFT。BIFT 生成过程如下：

a. BFR 通过 IGP 协议将 BIER 层配置的 BIER 信息进行通告。

b. BFR 基于 IGP 协议通告的 BIER 信息，在 BFR 之间的最佳转发路径上生成 BIFT。

在转发平面，当封装有 BIER 头的组播报文在 BIER 层进行转发时，BFR 根据 BIER 头中的信息查找 BIFT 表项完成报文复制转发。

(3) Overlay 层

Overlay 层在控制平面主要负责组播业务控制面信息交互，比如 BFIR 和 BFER 之间用户组播的加入和离开信息，建立组播流与 BFER 对应关系的组播转发表项。

在转发平面，当组播报文到达 BFIR 时确定目的 BFER 集合，并完成该组播报文对应的 BIER 头的封装；当携带有 BIER 头的组播报文到达 BFER 节点时，解封装 BIER 头并完成后续的组播报文转发。

5.5.5 报文封装格式

RFC 8279 中对 BIER 定义了多种封装类型，不同的封装类型报文格式不相同。目前，我司支持 G-BIER 和 BIERv6 封装类型。

- G-BIER（Generalized BIER，通用位索引显式复制）为中国移动携手新华三、华为、中兴等厂商提出的一种通用 BIER 封装方案，它根据 IPv6 网络的特点对 RFC 定义的标准的 BIER 头进行适配性修改，与 IPv6 实现了更好的融合。
- BIERv6（Bit Index Explicit Replication IPv6 Encapsulation，IPv6 封装的比特索引显式复制）是基于 Native IPv6 的全新组播方案，BIERv6 结合了 IPv6 和 BIER 的优势，可以无缝融入 SRv6 网络，简化了协议复杂度。将 BIER 承载报文的封装类型配置为 BIERv6 时，需要 BIER 子域内的所有的 BFR 均支持 SRv6。

1. G-BIER 报文封装格式

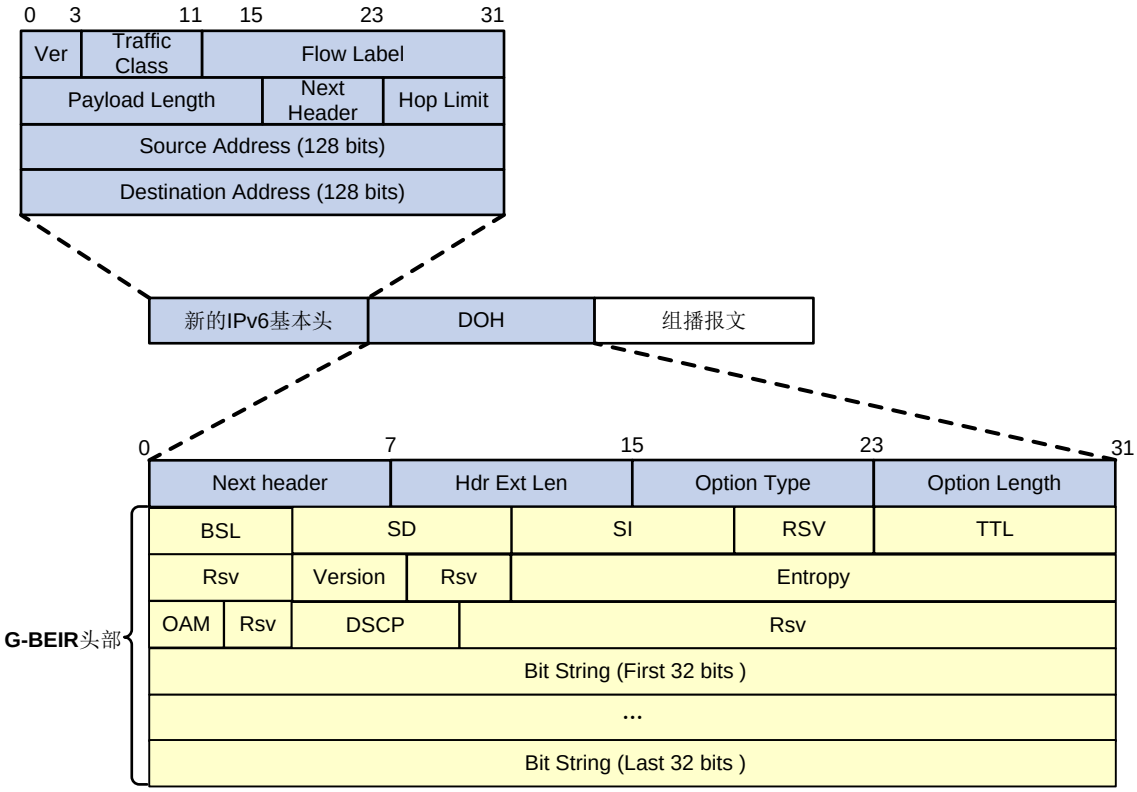
对 G-BIER 报文的封装是通过在组播数据报文前面添加新的 IPv6 基本头和 G-BIER 头实现的。如图 62 所示，IPv6 基本头中 Next Header 取值为 60，表明下一个报文头为 DOH（Destination Options Header，目的选项头）。

对于 G-BIER 报文，IPv6 基本头中有如下约定：

- Source Address：源地址需要配置为 BFIR 的组播服务源地址，该源地址由 BFIR 的前缀地址和组播服务 ID 值共同生成。BFIR 的前缀地址用来标识 BFIR 的网络位置，组播服务 ID 用来标识不同的 MVPN 实例。组播报文在转发过程中，该源地址保持不变。

- Destination Address：目的地址需要配置为专门用于 BIER 转发的 MPRA（Multicast Policy Reserved Address，组播策略保留地址），该地址要求在子域内路由可达。当 BFR 收到 IPv6 报文中的目的地址为本地配置 MPRA，则表示需要对该报文进行 G-BIER 转发。

图62 G-BIER 的报文封装示意图



G-BIER 报文中的 BIER 头主要包含以下几个部分：

- Next Header：8bits，用来标识下一个报文头的类型。
- Hdr Ext Len：8bits，表示 IPv6 扩展头长度。
- Option Type：8bits，选项类型为 G-BIER。
- Option Length：8bits，选项长度。
- BSL：4bits，取值用 1～7 来代表不同比特串长度，取值与比特串长度的对应关系如下：
  - 1：表示比特串长度为 64bits。
  - 2：表示比特串长度为 128bits。
  - 3：表示比特串长度为 256bits。
  - 4：表示比特串长度为 512bits。
  - 5：表示比特串长度为 1024bits。
  - 6：表示比特串长度为 2048bits。
  - 7：表示比特串长度为 4096bits。
- SD：8bits，BIER 子域 ID。
- SI：8bits，集标识。
- Rsv：保留字段。

- TTL：8bits，和 IP 报文中的 TTL 意义相同，可以用来防止环路。
- Version：4bits，版本号，目前只支持 0。
- Entropy：20bits，用来在存在等价路径时，进行路径的选择。拥有相同 Bit String 和 Entropy 值的报文，选择同一条路径。
- OAM：4bits，缺省值为 0，可用于 OAM 功能。
- DSCP：7bits，报文自身的优先等级，决定报文传输的优先程度。
- Bit String：位串。

2. BIERv6 报文封装格式

对 BIERv6 报文的封装是通过在组播数据报文前面添加新的 IPv6 基本头和 BIERv6 头实现的。如图 63 所示，IPv6 基本头中 Next Header 取值为 60，表明下一个报文头为 DOH（Destination Options Header，目的选项头）。

对于 BIERv6 报文，IPv6 基本头中有如下约定：

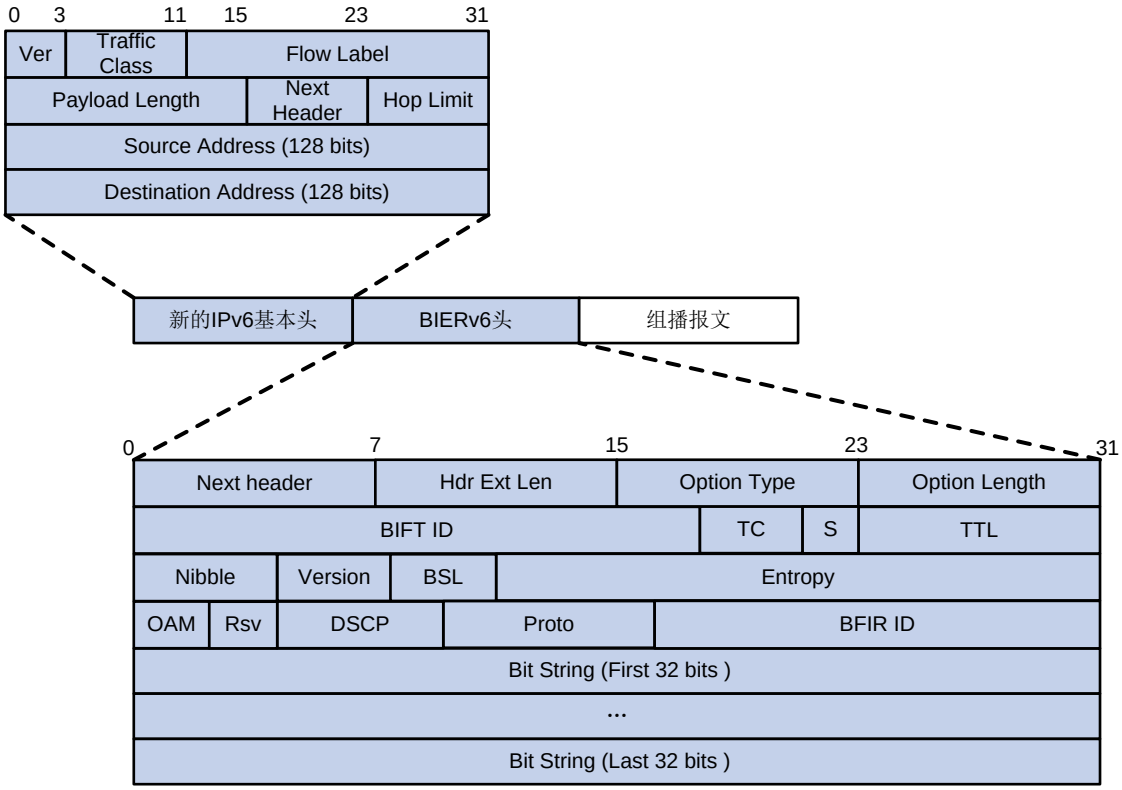
- Source Address：源地址需要配置为 BIERv6 隧道的源地址，在组播报文在公网中转发时，该源地址保持不变。有关 BIERv6 隧道源地址的详细介绍，请参见“IP 组播配置指导”中的“组播 VPN”。
- Destination Address：目的地址需要配置为专门用于 BIER 转发的 End.BIER SID，该地址要求在子域内路由可达，可通过 **`end-bier locator`** 命令进行配置。

BIERv6 头主要包含以下几个部分：

- Next Header：8bits，用来标识下一个报文头的类型。
- Hdr Ext Len：8bits，表示 IPv6 扩展头长度。
- Option Type：8bits，选项类型为 BIERv6。
- Option Length：8bits，表示 BIERv6 报文头长度。
- BIFT-ID：20bits，位索引转发表 ID，用来唯一标识一张 BIFT。
- TC：3bits，Traffic Class，流量等级，用于 QoS。。
- S：1bits，可视为保留字段。
- TTL：8bits，表示报文经过 BIERv6 转发处理的跳数。每经过一个 BIERv6 转发节点后，TTL 值减 1。当 TTL 为 0 时，报文被丢弃。
- Nibble：4bits，保留字段，目前只支持 0。
- Version：4bits，BIERv6 报文版本号，目前只支持 0。
- BSL：4bits，取值用 1～7 来代表不同比特串长度，取值与比特串长度的对应关系如下：
  - 1：表示比特串长度为 64bits。
  - 2：表示比特串长度为 128bits。
  - 3：表示比特串长度为 256bits。
  - 4：表示比特串长度为 512bits。
  - 5：表示比特串长度为 1024bits。
  - 6：表示比特串长度为 2048bits。
  - 7：表示比特串长度为 4096bits。
- Entropy：20bits，用来在存在等价路径时，进行路径的选择。拥有相同 Bit String 和 Entropy 值的报文，选择同一条路径。

- OAM：2bits，缺省为 0，可用于 OAM 功能。
- Rsv：2bits，保留字段，缺省为 0。
- DSCP：6bits，报文自身的优先等级，决定报文传输的优先程度。
- Proto：6bits，下一层协议标识，用于标识 BIERv6 报文头后面的 Payload 类型。
- BFIR ID：16bits，BFIR 的 BFR ID 值。

图63 BIERv6 的报文封装示意图

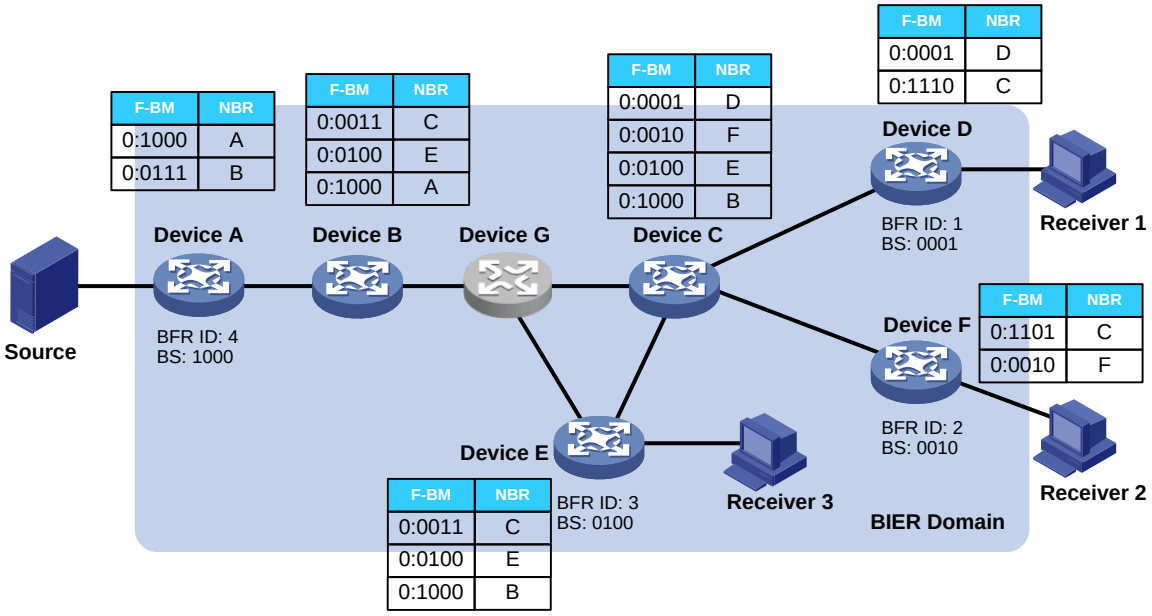


5.5.6 BIER 控制平面

在 BFR 上配置的 BIER 信息（SD、BFR prefix、BFR ID 等），通过 IGP 协议在 BIER 域内泛洪。IGP 根据邻居泛洪的 BIER 信息计算 BIER 最短路径树（以 BFIR 为根，Transit BFR 和 BFER 为叶子）。BFR 根据 BIER 最短路径树，生成 BIRT，最终进一步生成用于指导 BIER 转发的 BIFT。

如图 64 所示，BIER 域中包含支持 BIER 和不支持 BIER 的节点（Device G）。对于不支持 BIER 的节点，会将其所有的子节点作为叶子节点加入到 BIER 最短路径树，如果子节点不支持 BIER，则继续往下迭代到支持 BIER 的子节点。比如，Device G 不支持 BIER，Device G 会将子节点 Device C 和 Device E 的 BIER 信息传递给 Device B，在 Device B 上生成下一跳邻居为非直连邻居 Device C 和 Device E 的 BIFT 表项信息。

图64 BIER 控制平面 BIFT



5.5.7 BIER 转发过程

组播报文到达 BFIR，BFIR 查找组播转发表得到该组播表项对应的 BIFT-ID 和 BS，BS 即组播报文穿越 BIER 域后到达的全部 BFER 集合。BFIR 根据 BIFT-ID 匹配到指定 BIFT，并根据报文头中携带的 Bit String 与 BIFT 表项匹配计算后复制转发。BIER 报文到达 BFER 节点后解封成组播报文，根据组播地址查找组播转发表进行转发。

当 BIER 转发过程中需要经过非 BIER 节点，即 BFR 的下一跳邻居为非直连邻居时，可以通过特定的技术来穿越非 BIER 节点。特定的技术取决于 BIER 封装的外层封装（比如，MPLS 封装依靠 LSP 穿过非 BIER 节点，IPv6 封装可以按普通 IPv6 单播路由到非直连 BIER 邻居）。

下面以具体的示例来说明 BIER 转发过程。如图 65 所示，BIER 子域中的每台设备都根据 IGP 协议计算生成了 BIFT。Device D 和 Device E 的下游存在某个组播组的接收者，Device A 作为 BFIR 通过 BGP MVPN 路由收集到 Device D 和 Device E 上与 Device A 处于相同 BIER 子域的 MVPN 信息。当 Device A 收到发往该组播组的组播报文时，BIER 转发过程如下：

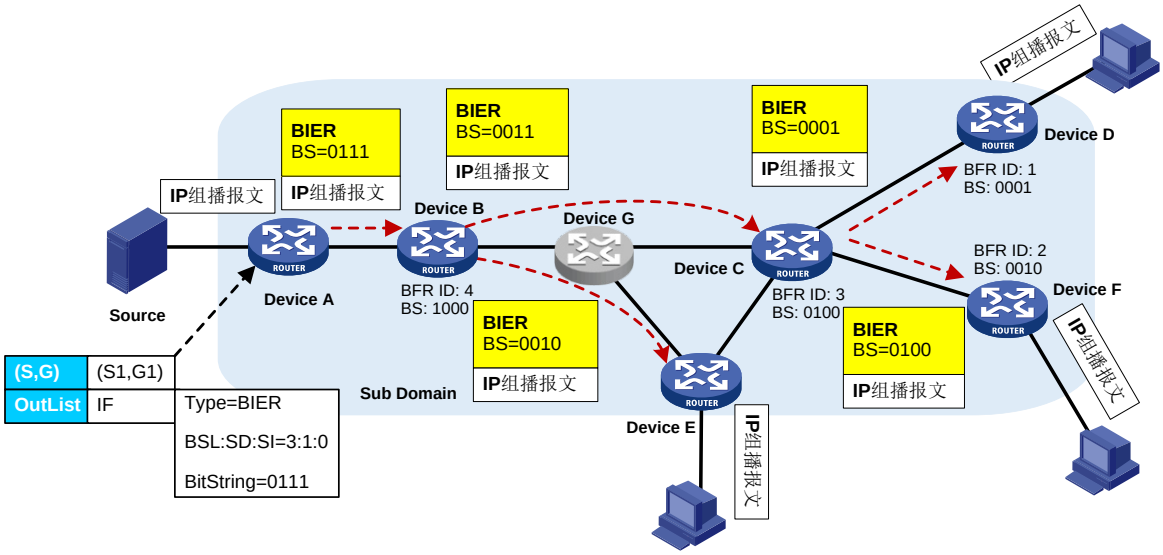
(1) Device A 收到 IP 组播报文后，查找组播转发表项，得到该组播表项对应的 BIFT-ID 和 BS，根据 BIFT-ID 找到对应的 BIFT 表，将 BS 与 BIFT 表中每行表项 F-BM 进行“位与”操作，复制组播报文并按照 BIER 报文格式封装（封装的 BS 为“位与”计算后得到的值），发送给下一跳邻居 Device B。

(2) Device B 收到 BIER 报文后，根据 BIER 头中的 BIFT-ID 和 BS，执行与步骤(1)相同的步骤，发现下一跳邻居为 Device C 和 Device E，需要经过非 BIER 的节点 Device G。此时，可以通过特定的技术穿越非 BIER 节点，复制组播报文并按照 BIER 报文格式封装后，发给 Device C 和 Device E。

(3) Device C 收到 BIER 报文后，执行与步骤(1)相同的操作，将组播报文复制一份，并按照 BIER 报文格式封装后，发给 Device D 和 Device F。

(4) Device D、Device E 和 Device F 收到 BIER 报文后，发现只有本节点对应的 F-BM 与上游发送的 BIER 报文中的 BS 进行“位与”操作后不为 0，表明本节点为 BFER，需要结束 BIER 转发。此时，Device D、Device E 和 Device F 分别从 BIER 头部解封装出组播报文后，根据组播路由表项继续转发给下游接收者。

图65 BIER 报文转发过程示意图



注：为了简化示意**BIER**转发过程，此处**BS**长度使用**4bits**进行演示，实际取值以设备支持情况为准

6 IPv6 部署方案

6.1 IPv6升级改造方案

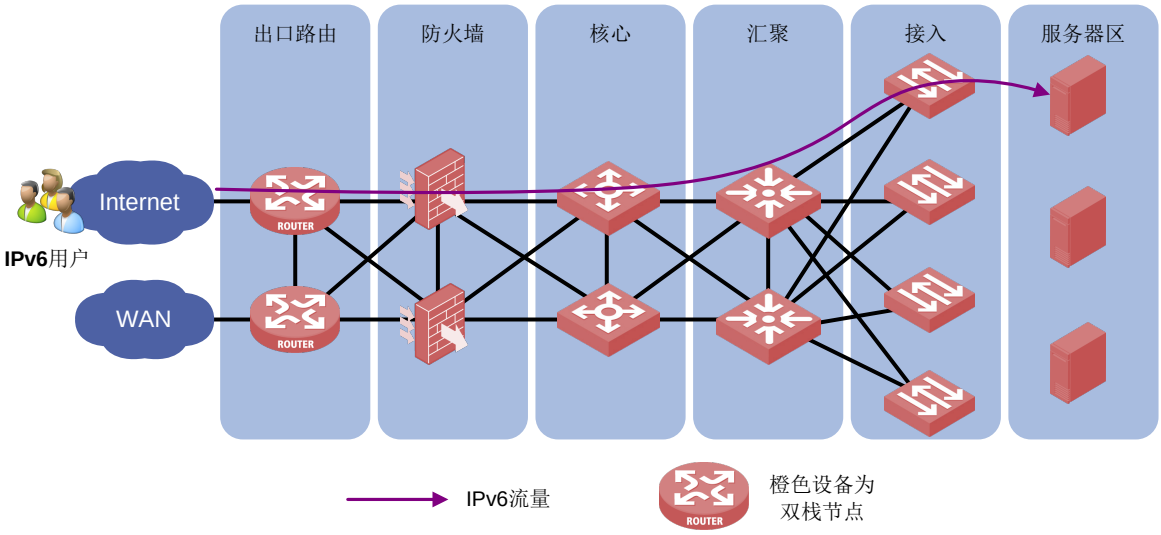
IPv6 升级改造方案包括如下几种：

- 新建 IPv6 网络
- 部分设备支持双栈
- 网络边界进行地址翻译

6.1.1 新建 IPv6 网络

新建 IPv6 网络方案是指按照现有的网络建设模式组建全新的 IPv6 网络，网络中的设备均支持 IPv4 和 IPv6 双栈，该网络仅用来处理 IPv6 流量。全新的 IPv6 网络中可以部署新业务，以实现业务创新。

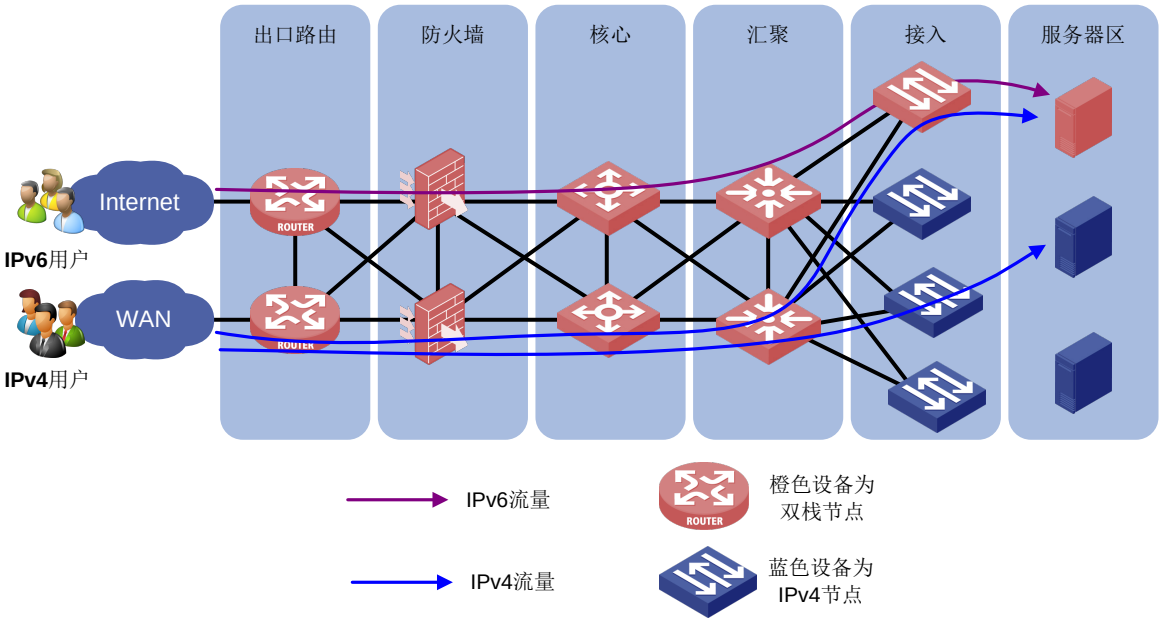
图66 新建 IPv6 网络示意图



6.1.2 部分设备支持双栈

新建 IPv6 网络方案是指仅网络出口、防火墙、核心设备、部分汇聚/接入设备支持双栈，IPv6 用户通过这些双栈节点进行通信。

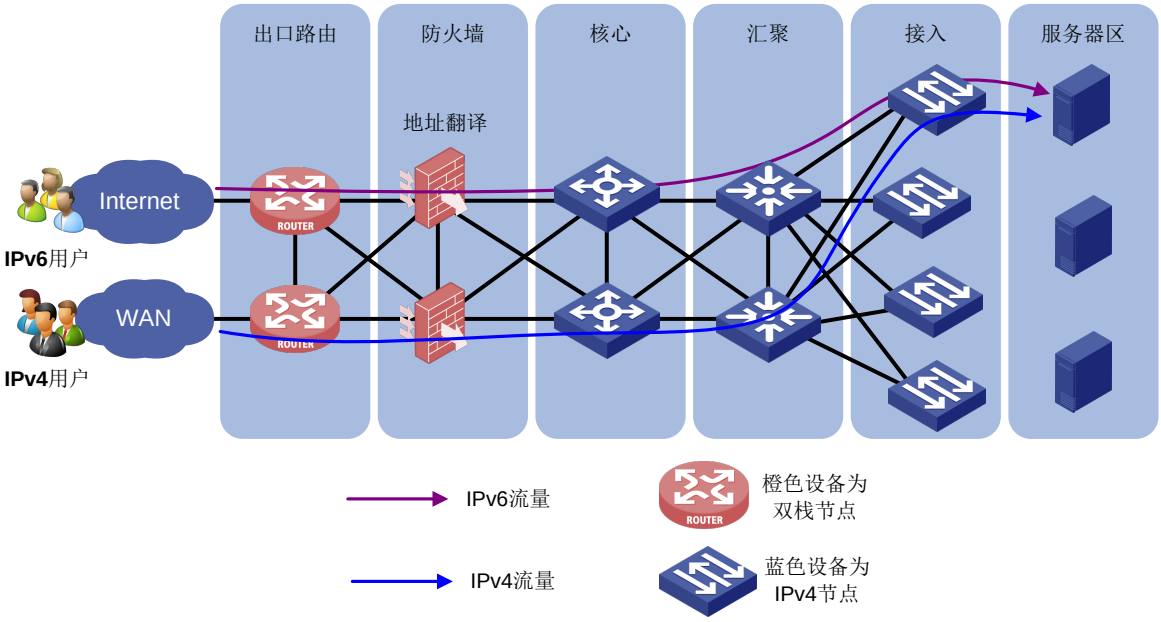
图67 部分设备支持双栈示意图



6.1.3 网络边界进行地址翻译

网络边界进行地址翻译方案是指在位于 IPv4 与 IPv6 网络边界的防火墙上开启 AFT 等地址转换协议，对 IPv4 和 IPv6 地址进行相互转换，从而实现 IPv6 用户访问 IPv4 网络。

图68 网络边界进行地址翻译示意图



6.1.4 升级改造方案对比

IPv6 网络的三种升级改造方案对比如所示。

表11 升级改造方案对比表

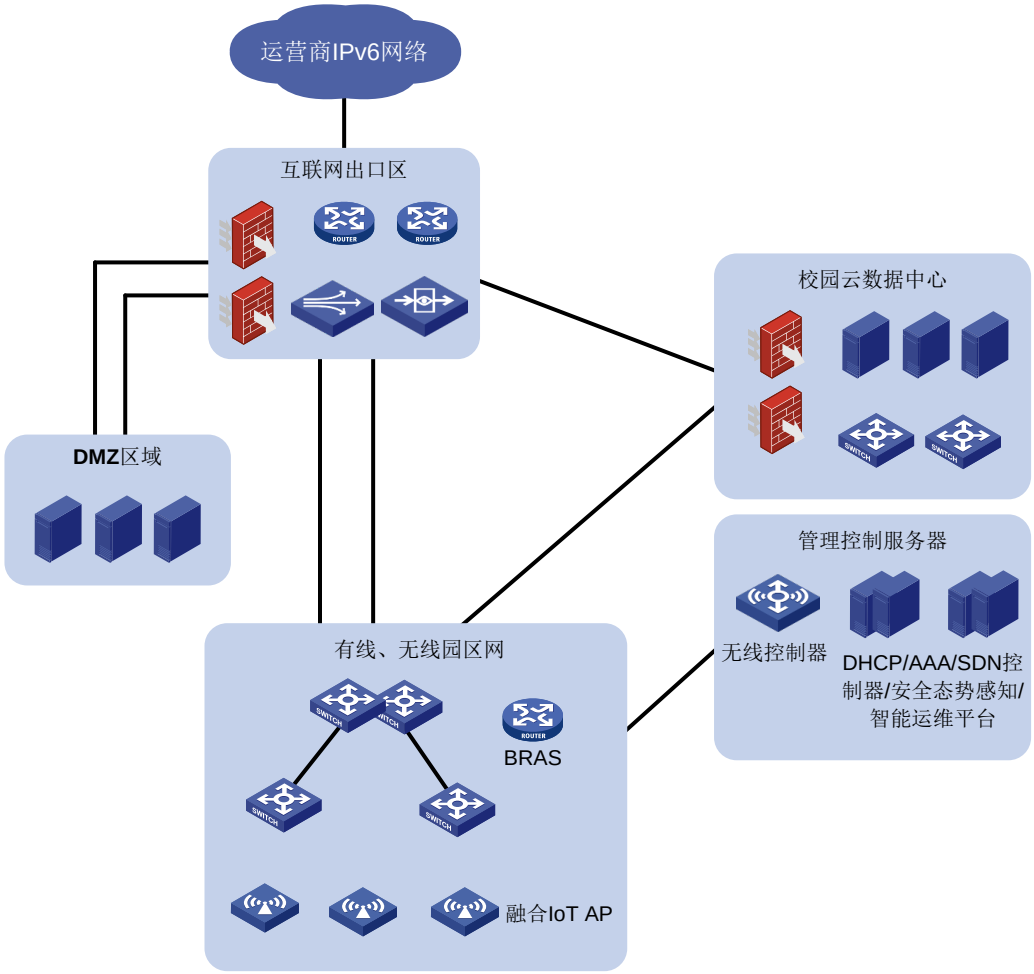
升级改造方案	新建 IPv6 网络	部分设备支持双栈	网络边界进行地址翻译
改造成本	高	中	低
改造难度	低	高	低
改造工作量	中	中	低
改造风险	低	高	中
对原有业务影响	无影响，IPv6业务完全重新部署	有影响，IPv4和IPv6应用复用双栈网络	无影响，原有IPv4业务不需要进行地址转换
单点设备压力	低 IPv4和IPv6两张网络，由不同的设备处理IPv4和IPv6流量	中 网络、防火墙等各节点需部署双栈协议，设备资源表项共享	高 地址转换的节点成为网络瓶颈
运维复杂度	IPv4/IPv6运维界面分离，运维简单	IPv4/IPv6运维界面混合，运维复杂	延续IPv4运维方式，运维简单
适用场景	业务系统重要、架构复杂，改造难度大的网络（如金融网银业务等） 具有业务创新需求的网络	结构清晰、简单的网络	资金预算紧张，或希望尽量减少IPv6对现网冲击的网络

6.2 园区网全面IPv6化部署方案

园区网全面 IPv6 化的典型部署方案如图 69 所示。具体部署方法为：

- 在互联网出口区：
  - 防火墙、IPS、LB、应用控制、流量分析等设备全面支持 IPv6。
  - 全面支持 AFT 等地址转换协议。
- 在校园云数据中心：
  - Underlay、Overlay 网络支持 IPv6。
  - 虚拟化、云平台全面支持 IPv6。
  - 部署 AFT 等地址转换协议。
- 在管理控制服务器区：
  - 安全态势感知支持 IPv6。
  - 认证授权、网管服务器支持 IPv6。
- 在有线、无线园区网：
  - 支持 IPv6 地址分配，IPv6 路由协议。
  - 支持 IPv6 准入（BRAS 或 ADCampus）。
  - 支持有线、无线接入 SAVA 和域内 SAVA，防止源地址假冒攻击。
- 在 DMZ 区域：设置 DNS64、对外 IPv6 化网站群等。

图69 园区网全面 IPv6 化典型组网

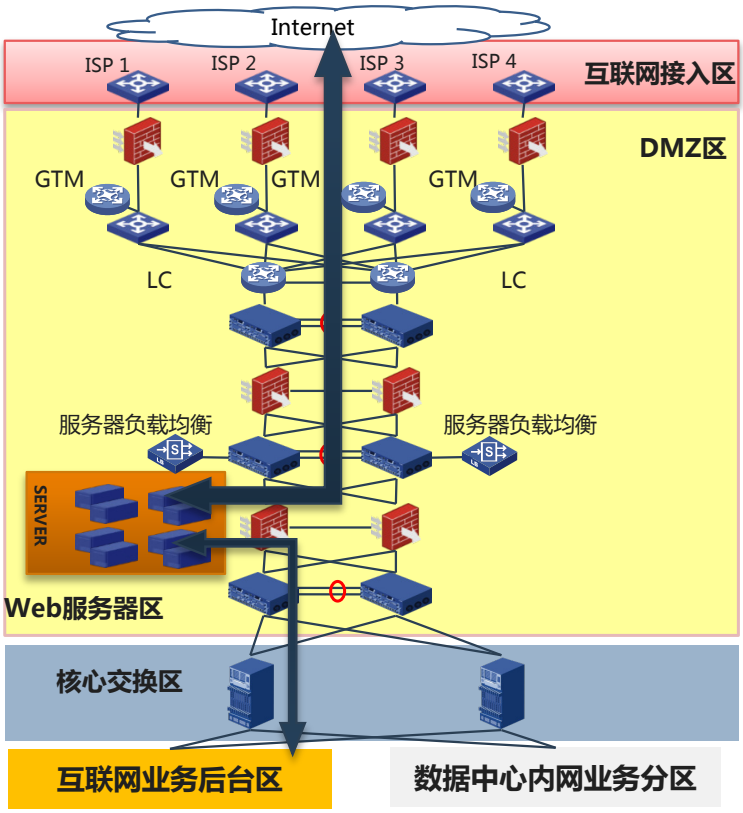


6.3 金融网络IPv6改造方案

如图 70 所示，金融网络由以下几部分组成：

- 互联网接入区：通过 ISP 的 Internet 线路完成用户的网络接入功能，该区域部署广域网接入路由器，实现多 ISP 的多条 Internet 线路接入。
- DMZ（Demilitarized Zone，隔离区）区：为用户提供 Web 服务的服务器位于 DMZ 区。通过防火墙和 IPS，实现互联网与 DMZ 区隔离；DMZ 区内部署路由器和交换机，实现本区域内所有设备的互联互通；通过负载均衡设备，优化业务响应速度并保证 Web 业务高可用性。
- 核心交换区：连接 DMZ 区和互联网业务后台区、数据中心内网服务器区。
- 互联网业务后台区：主要提供门户、网银、互联网金融业务的应用处理。
- 数据中心内网服务器区：包含多个业务区，主要为 APP 服务器或 DB 提供数据服务。

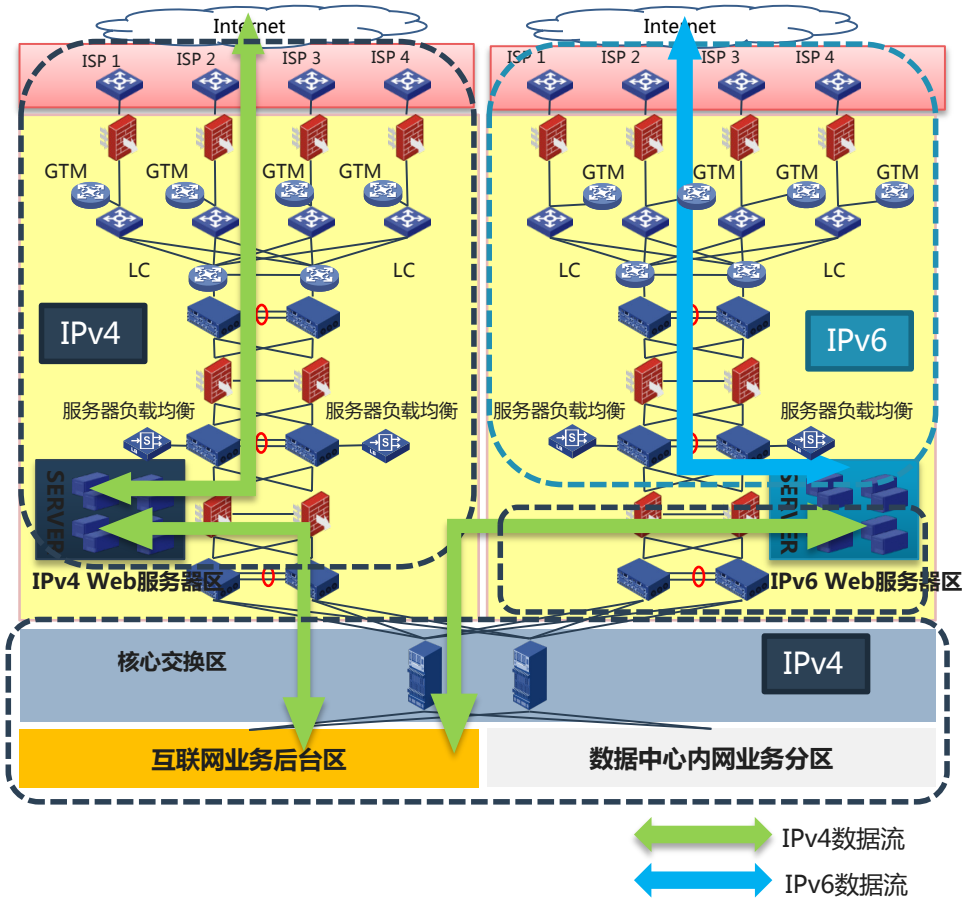
图70 金融网络架构示意图



可以通过以下几种方案实现金融网络从 IPv4 到 IPv6 网络的升级：

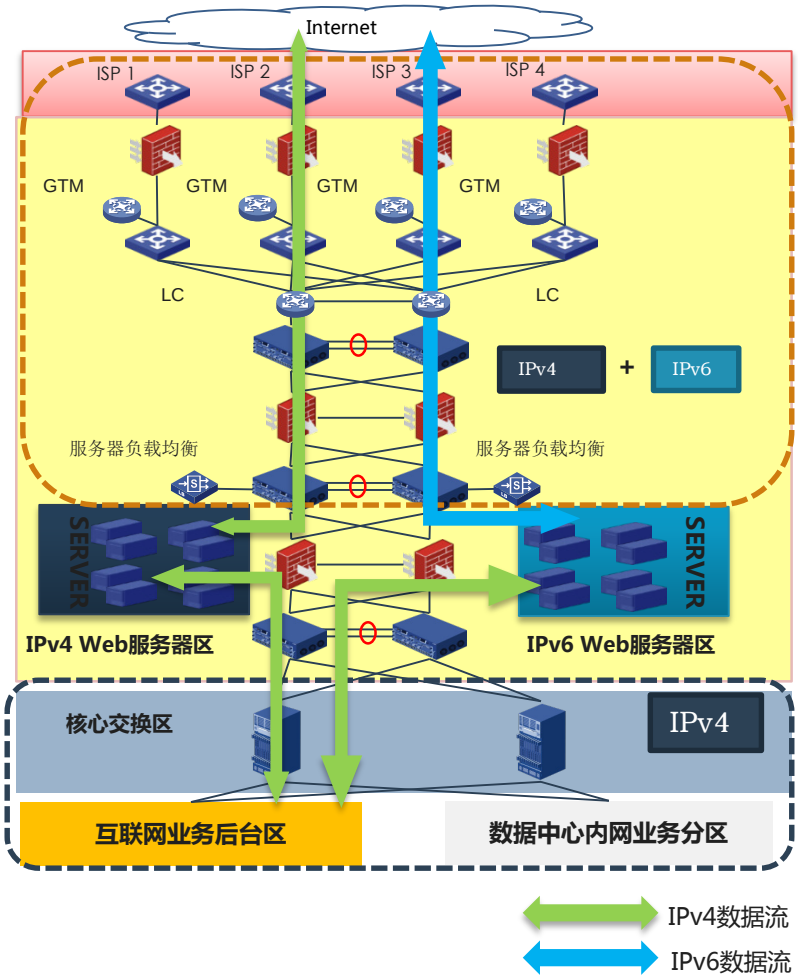
- 完全新建纯 IPv6 区：新建纯 IPv6 的互联网接入区和 DMZ 区。互联网接入区和 DMZ 区通过 IPv6 对外提供 Web 服务，通过 IPv4 和后端业务后台区、内网业务区数据拉通

图71 完全新建纯 IPv6 区



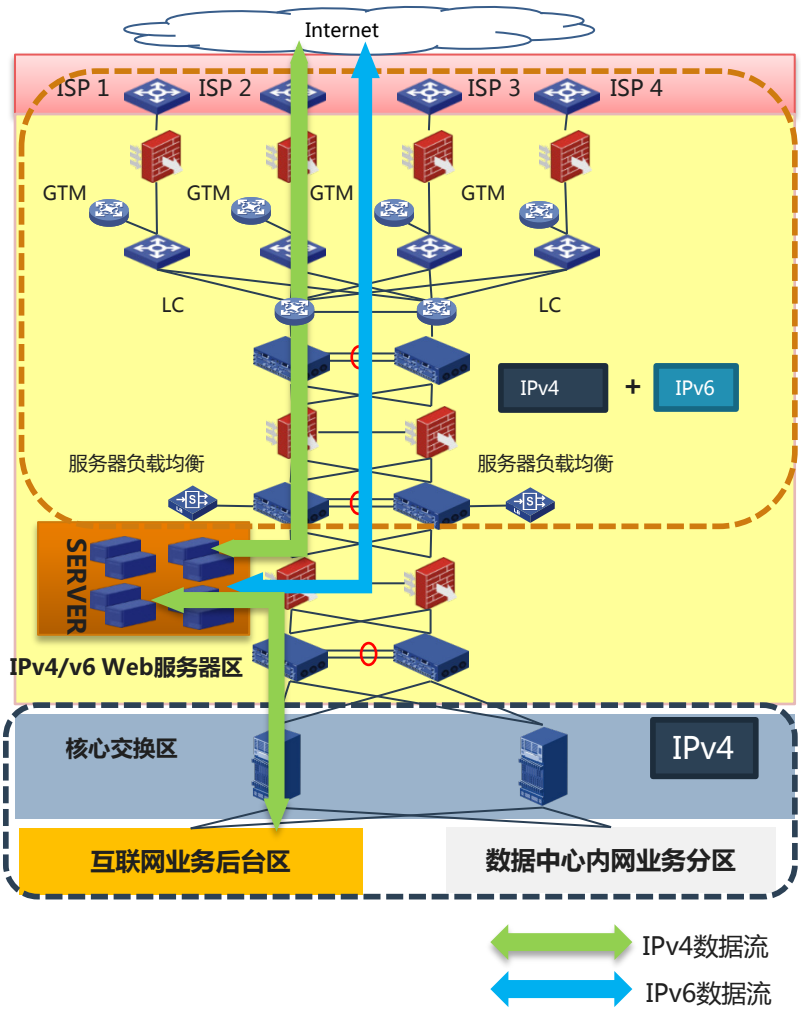
- 新建 IPv6 Web 服务器集群：改造互联网接入区和 DMZ 区支持双栈，原有 IPv4 Web 服务器集群和新建的 IPv6 Web 服务器集群共用双栈网络。

图72 新建 IPv6 Web 服务器集群



- 原 Web 服务器集群网络升级为双栈网络：改造互联网接入区、DMZ 区和 Web 服务器集群网络支持双栈。

图73 原 Web 服务器集群网络升级为双栈网络



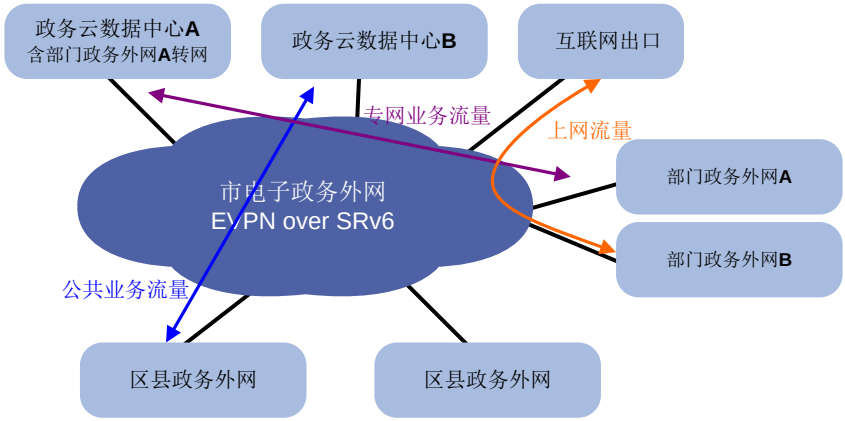
6.4 电子政务外网IPv6+应用

6.4.1 SRv6 应用

电子政务外网中 SRv6 的部署方案如图 74 所示。

- 访问互联网的上网流量通过 EVPN over SRv6 承载，为上网用户分配相应 VPN 权限，并通过 SRv6 隧道进行流量承载。
- 公共业务流量可以使用公网地址，直接通过 SRv6 进行承载。
- 专网流量进行相应 EVPN 的划分，保证专网专用，业务逻辑隔离。
- 所有路径都可以由 SDN 控制器进行动态调整，保证资源利用率最优。

图74 电子政务外网的 SRv6 应用

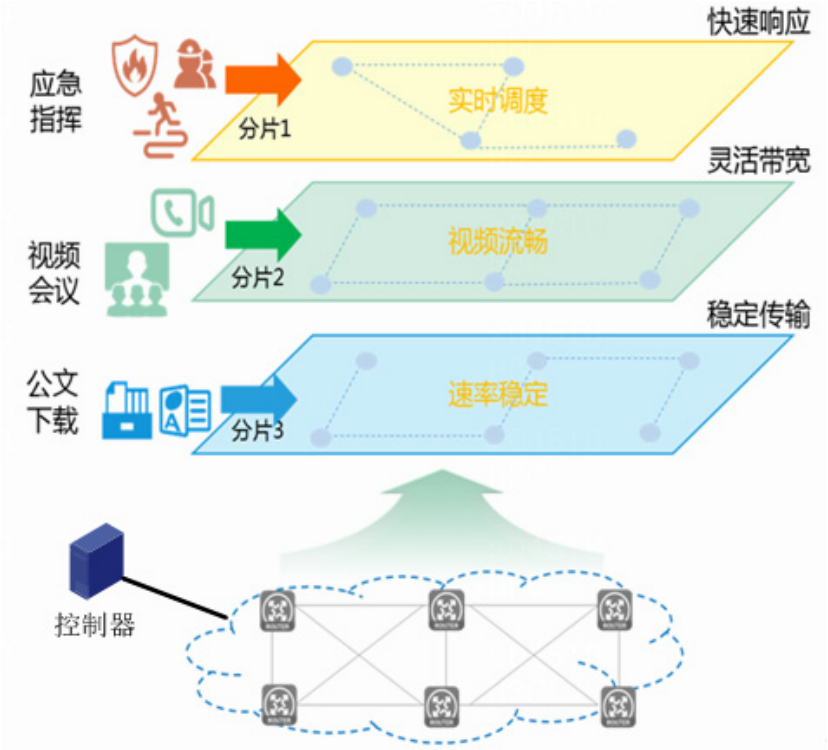


6.4.2 网络切片应用

电子政务外网中网络切片的部署方案如图 75 所示。

- 综合运用多种网络切片技术，将骨干网带宽细化，形成多通道。
- 根据不同的业务需求（时延/抖动/丢包等），通过控制器进行不同策略的下发，实现对通道的多样性利用。
- 可以根据需求为不同专网（例如应急指挥网络、视频会议网络、公文下载网络等）分配不同的带宽，保障在同一拓扑中，一个专网的流量不会因为另一个专网流量的拥塞而丢包。
- 可根据客户的业务需求灵活编排 SRv6 转发路径，提升网络智能性。

图75 电子政务外网的网络切片应用



6.4.3 可视化应用

电子政务外网中可视化的部署方案如图 76 所示。通过基于 iFIT 和 Telemetry 技术的可视化方案可以实现：

- 质量可视，规划支撑：周期性的数据收集，形成现网报表数据，为扩容及后续规划等提供数据支撑。
- 随需而动，智能运维：根据现网业务状态，实现网络路径等智能调优，保障关键业务质量。
- 精准定位、快速排障：业务出现问题时，通过网络分析器上的图形展示可快速定界和解决问题。

说明

Telemetry 的详细介绍，请参见《Telemetry 技术白皮书》。

图76 电子政务外网的可视化应用