江苏省邮电规划设计院有限责任公司专家团队

精品力作

下一代互联网IPv6过渡技术与部署实例

NEXT GENERATION INTERNET IPv6
TRANSITION TECHNOLOGY AND DEPLOYMENT WITH INSTANCES

戴 源 杨 建 袁 源 李忠超 石启良 张家燕 编著

人民邮电出版社
POSTS & TELECOM PRESS

下一代互联网IPv6过渡技术与部署实例

封面设计：王娟

人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-33343-8

ISBN 978-7-115-33343-8

定价：45.00 元

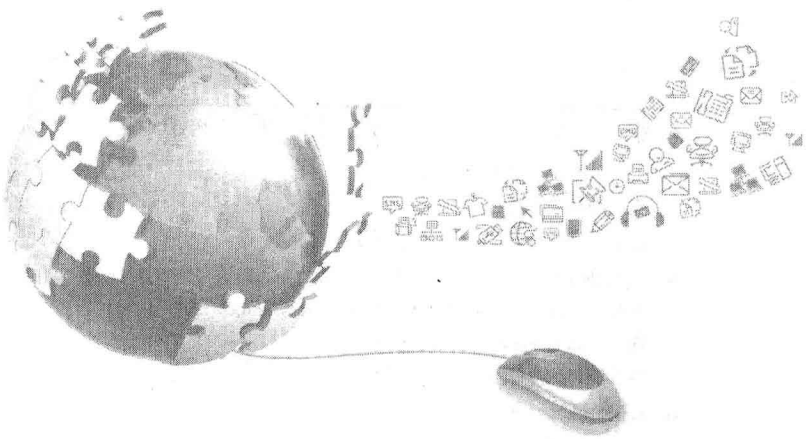
江苏省邮电规划设计院有限责任公司专家团队

精品力作

下一代互联网IPv6过渡技术与部署实例

NEXT GENERATION INTERNET IPv6
TRANSITION TECHNOLOGY AND DEPLOYMENT WITH INSTANCES

戴 源 杨 建 袁 源 李忠超 石启良 张家燕 编著



人民邮电出版社

图书在版编目（CIP）数据

下一代互联网IPv6过渡技术与部署实例 / 戴源等编著. -- 北京 : 人民邮电出版社, 2014.1
ISBN 978-7-115-33343-8

Ⅰ. ①下… Ⅱ. ①戴… Ⅲ. ①计算机网络－通信协议－研究 Ⅳ. ①TN915.04

中国版本图书馆CIP数据核字(2013)第238670号

内 容 提 要

本书重点描述了IPv6过渡技术和方案，然后结合实际情况介绍了网络、应用、终端的部署方案和策略，最后给出了两个网络部署实例和一个IPTV业务系统IPv6改造实例。全书分为三部分：第一部分全面介绍了IPv6的基础知识；第二部分描述了IPv6过渡技术方案，对主流的转换技术双栈、DS-Lite、NAT64技术做了对比分析，然后结合运营商实际讲述了过渡技术的部署方案，最后介绍了应用和终端的具体过渡方案；第三部分给出了两个实验部署实例，这部分有助于加深对前面章节的理解，对准备参与IPv6建设的数据人员也是一个较好的实际参考。

本书内容全面，除理论介绍外，书中多处结合了作者参加IPv6试点项目的实践经验，阐述了实验中需要解决的各类问题，工程实践性较强，可供从事IP网络相关工作的工程技术人员使用，也可供高等院校计算机、通信、网络等专业的师生阅读参考。

◆ 编　著　戴　源　杨　建　袁　源　李忠超　石启良
张家燕
责任编辑　杨　凌
责任印制　彭志环　焦志炜

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路11号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
三河市海波印务有限公司印刷

◆ 开本：787×1092　1/16
印张：13.5
字数：321千字　　2014年1月第1版
印数：1－4 000册　　2014年1月河北第1次印刷

定价：45.00元

读者服务热线：(010) 81055410　印装质量热线：(010) 81055316
反盗版热线：(010) 81055315

序

在全球互联网快速发展的时期，中国的互联网市场也有了突飞猛进的发展，用户数和上网时长随着网络规模的扩大不断增加。互联网通信能力的快速提升也推动了各项网络应用的发展。随着物联网等新兴业务应用的发展，将会使所有的信息设备智能化联网，IP 地址的需求将越来越大。在过去的两年中，全球各大地址机构陆续宣布 IPv4 地址已经分配完毕，因 IP 地址不足制约互联网蓬勃发展的矛盾日益突出。

IPv6——这个被设计用来取代 IPv4 的下一代互联网协议，已不仅停留在技术研究和标准制定阶段，启动以 IPv6 为基础的下一代互联网部署已经刻不容缓。然而，下一代互联网部署涉及互联网技术、网络设备、用户终端、操作系统、内容及应用等众多产业链环节，向下一代互联网的过渡将是一个长期的系统工程，不能一蹴而就。

我国政府一直高度重视下一代互联网的发展和演进，早在 2004 年就启动了中国下一代互联网示范工程（CNGI）建设，2012 年下发的《下一代互联网“十二五”发展建设意见》，更是吹响了我国下一代互联网规模部署实施的号角。而要实现下一代互联网规模商用的整体目标，必然依赖于产业链各方的共同努力。对于运营商而言，向下一代互联网的部署不只是网络上 TCP/IP 协议栈的简单升级，关键还在于如何平稳实现当前规模庞大的互联网用户和应用的迁移，确保迁移过程中服务质量和用户体验不受影响，并满足各类新型业务发展的需求。

中国电信作为国内最大的宽带运营商，多年来在下一代互联网领域进行了大量技术研究，并在江苏、湖南等省开展了现网试点，探索网络部署方案和业务运营模式。本书作者即为参与 IPv6 项目试点的核心成员，在运营商网络规划、设计过程中积累了大量的有益经验。

本书聚焦于运营商下一代互联网的网络部署和演进，特点是理论和实践相结合，有较强的实用性与前瞻性，既有 IPv6 基础知识和过渡技术的介绍，也有运营商网络部署的业务迁移方案；既有具体的设备配置实例，也有终端演进方面的思考。书中内容很多是作者的现网实践经验，相信会给广大读者加深对下一代互联网的理解提供帮助。

2013 年 11 月

前　言

在全球互联网快速发展的时期，中国的互联网市场也有了突飞猛进的发展，用户数和上网时长随着网络规模的扩大也不断增加。我国互联网通信能力的快速提升，带动了我国宽带互联网的发展，宽带的快速普及推动了各项网络应用的发展。互联网及其应用快速发展，随之而来的是 IPv4 地址的大量消耗，IANA 的 IPv4 地址在 2011 年 2 月 3 日已经分配完毕，5 个 RIR 管理的地址空间中，RIR 亚太地区已经分配完毕。而随着物联网的发展，所有的信息设备将会实现智能化联网，对 IP 地址的需求将越来越大。

业界早已认识到上述问题的严峻性，并且开展了一系列研究。在这种环境下，IPv6 应运而生。1992 年 IETF 成立工作组规划下一代互联网；1994 年 IETF 会议决定正式提议 IPv6 发展计划；1995 年 IETF 发布 IPv6 第一个正式 RFC——RFC 1883；1998 年 IETF 正式发布 IPv6 标准规范——RFC 2460；2000 年以后周边协议开始增加对 IPv6 的支持；2005 年以后运营商关注过渡问题，过渡技术研究日益增多。

对比国际上其他国家，中国由于获得 IPv4 地址数量本身较少，而用户数又不断增多，需求显得尤为迫切。然而，由于没有商业利益的推动，网络运营商、应用开发商和信息设备制造商都不愿意大力推动。内容源匮乏、终端支持力度较弱、国家安全监管存在漏洞等因素导致互联网业务全程 IPv6 承载进展缓慢，必须依靠政府的引导来加快推进。前国务院总理温家宝 2011 年 12 月 2 日主持召开国务院常务会议，研究部署加快发展我国下一代互联网产业；2012 年 2 月 16 日国家发改委办公厅发布“关于组织实施 2012 年下一代互联网技术研发、产业化和规模商用专项的通知”；2012 年 3 月 30 日国家七部委联合发布“关于印发下一代互联网（“十二五”）发展建设的意见”。

在政府的大力推动下，各大运营商都出台了一系列具体措施予以落实并积极参与下一代互联网发展工作，以促进产业链的发展。本书作者即为协助大型网络运营商设计、规划 IPv6 的网络演进，参与 IPv6 项目试点的成员。在此项工作过程中作者也积累了一定的理论和实践经验。在查阅文献书籍的过程中，作者发现现有关于 IPv6 技术的书籍分为两类：一类侧重于理论和实验，另一类侧重于向 IPv6 过渡过程中各部分的具体方案。但是，既结合了理论和方案，又给出了具体实验场景、具体设备配置参考的书籍还不是很多，因此编纂此书，希望在普及 IPv6 基础知识、过渡技术的基础上，加入这些年参加 IPv6 项目试点的经验总结，以期给广大从事下一代互联网发展的 IP 从业人员一个实施方案和具体细节的参考。同时也可以让广大高等院校相关专业学生和网络技术爱好者在学习 IPv6 基础知识、过渡技术的同时，

结合实践方案以加深理解。

全书分为三部分，共 9 章，各个部分的内容介绍如下。

第一部分：包括第 1 章到第 3 章，这部分讲述 IPv6 基础知识。

第 1 章为 IPv6 简介，介绍了地址枯竭形势和应对手段，引出 IPv6 的发展历程和技术的新特点。第 2 章为 IPv6 技术基础，介绍了 IPv6 地址结构、ICMPv6 协议、ND 协议、DHCPv6 协议等基本协议以及 PPPoE 和 IPoE 这两种接入方式。第 3 章介绍 IPv6 多播技术，包括基础知识、多播地址、MLD 协议、IPv6 PIM 协议和部署策略。

第二部分：包括第 4 章到第 7 章，这部分首先介绍过渡方案，然后结合实际情况介绍网络、应用、终端的部署方案和策略。

第 4 章介绍 IPv6 过渡技术，主要有双栈、隧道、协议翻译技术等，然后对主流的转换技术双栈、DS-Lite、NAT64 技术做了对比分析，阐述了过渡技术部署的策略。第 5 章结合运营商实际讲述了过渡技术的部署方案，包括网络部署、业务部署、IPv6 地址规划、网络安全和业务溯源。第 6 章是内容和应用过渡的方案，从几个角度分类阐述，主要是 IDC、典型内容和应用（包括网页、FTP、邮件等常见应用）、大中型和小型 ICP、物联网等。第 7 章是终端演进的情况介绍，包括固网终端、无线终端、物联网终端等。

第三部分：包括第 8 章和第 9 章，第 8 章讲述了一个实验部署私网双栈和 DS-Lite 的实例，第 9 章讲述了一个实验部署 IPTV 业务系统 IPv6 过渡改造的实例。这部分有助于加深对前面章节的理解，对准备参与 IPv6 建设的数据人员也是一个较好的实际参考。

这部分结合了作者参加 IPv6 试点项目的实践经验，说明了实验的细节、流程、配置及实验中解决的各类问题，最后对实验方案做出小结。

本书由江苏电信下一代互联网推进办公室的主要成员执笔，主要由江苏电信资深专家杨建、李忠超以及江苏省邮电规划设计院戴源、袁源、石启良、张家燕、刘春林、王吉顺、胡波等多名专家和技术骨干联合编写。在书籍的编写过程中得到了华为技术有限公司和中兴通讯股份有限公司的大力支持，在此致以最衷心的感谢！由于编者水平所限，书中难免存在错误之处，欢迎各位读者批评指正！

编者

2013 年 8 月

目　录

第 1 章　IPv6 简介……1
1.1　IPv4 向 IPv6 过渡的必然性……1
1.1.1　国际国内 IPv4 地址枯竭形势……2
1.1.2　应对 IPv4 地址枯竭的手段……2
1.1.3　产业链各方向下一代互联网演进策略……3
1.2　IPv6 的发展历程及现状……4
1.3　IPv6 技术的新特点……5
1.3.1　地址管理……5
1.3.2　改进报头……5
1.3.3　改进路由……6
1.3.4　安全机制……7
1.3.5　增强服务质量……7

第 2 章　IPv6 技术基础……9
2.1　IPv6 地址介绍……9
2.1.1　IPv6 地址结构……9
2.1.2　IPv6 寻址模型……10
2.1.3　IPv6 地址分类……10
2.2　IPv6 报文结构……15
2.2.1　IPv6 基本报头……16
2.2.2　扩展分组头……16
2.3　ICMPv6 协议……22
2.3.1　ICMPv6 报文结构……22
2.3.2　ICMPv6 报文类型……22
2.4　邻居发现技术……23
2.4.1　ND 概述……24
2.4.2　ND 功能……28
2.5　IPv6 路由协议……31
2.5.1　IPv6 单播路由协议……31
2.5.2　IPv6 多播路由协议……34
2.6　IPv6 接入技术……35
2.6.1　DHCPv6……35
2.6.2　PPPoE……38
2.6.3　IPoEv6……40

第 3 章　IPv6 多播技术……42
3.1　IPv6 多播技术基础……42
3.1.1　多播的优点……42
3.1.2　多播分发树……43
3.1.3　反向路径转发……44
3.1.4　多播相关协议……44
3.2　IPv6 多播地址……44
3.2.1　IPv6 多播地址格式……45
3.2.2　IPv6 多播地址分配……45
3.2.3　永久分配的 IPv6 多播地址……45
3.2.4　基于单播前缀的多播地址……46
3.2.5　IPv6 SSM 多播地址……46
3.2.6　内嵌 RP 地址的多播地址……46
3.2.7　IPv6 多播 MAC 地址……47

3.3 MLD 协议 ……48
3.3.1 MLD 版本 ……48
3.3.2 MLD 原理简介 ……48
3.3.3 MLDv2 原理简介 ……50
3.4 IPv6 PIM 协议 ……52
3.4.1 PIM 消息格式 ……53
3.4.2 PIM 邻居发现和 DR 选举机制 ……53
3.4.3 PIM-SM 注册 ……56
3.4.4 PIM-SM 加入/剪枝 ……57
3.4.5 PIM-SM 最短路径树切换 ……59
3.4.6 PIM-DM 加入和剪枝过程 ……60
3.4.7 PIM-DM 嫁接机制 ……62
3.4.8 PIM 的断言机制 ……62
3.4.9 BSR 和 RP 选举 ……63
3.4.10 多路访问网络的处理 ……64
3.4.11 PIM-SSM ……64
3.5 IPv6 多播协议的现网部署策略建议 ……65
3.5.1 RP 部署 ……65
3.5.2 多播路由协议的部署 ……65
3.5.3 MLD 静态加入的部署 ……65
3.5.4 二层设备开启 MLD snooping ……66
3.5.5 RP 备份 ……66

第 4 章 IPv6 过渡技术 ……67
4.1 总体介绍 ……67
4.2 双栈技术 ……68
4.2.1 IP 骨干网 ……69
4.2.2 BNG ……69
4.2.3 CPE ……69
4.3 隧道技术 ……70
4.3.1 6PE ……70
4.3.2 6vPE ……71
4.3.3 6over4 隧道 ……71
4.4 协议翻译技术 ……73
4.4.1 NAT44 技术 ……73
4.4.2 DS-Lite 技术 ……82
4.4.3 NAT64 ……86
4.5 过渡技术分析、对比 ……87
4.6 过渡技术部署策略建议 ……87
4.6.1 部署双栈，构建网络向 IPv6 迁移的基石 ……87
4.6.2 部署 CGN，解决地址短缺，为业务平滑迁移提供保障 ……88
4.6.3 向纯 IPv6 网络过渡 ……89
4.6.4 全球运营商 IPv6 演进方案选择 ……90

第 5 章 运营商过渡技术部署 ……91
5.1 运营商网络演进策略总体分析 ……91
5.2 网络部署 ……92
5.2.1 网络架构和改造关键点分析 ……92
5.2.2 骨干网 ……94
5.2.3 城域网 ……97
5.2.4 接入网 ……102
5.2.5 IPv6 网络规划 ……105
5.3 业务部署 ……108
5.3.1 业务地址需求分析 ……108
5.3.2 宽带接入业务 ……109
5.3.3 无线宽带（含 3G 和 WLAN）业务 ……110
5.3.4 其他业务 ……111
5.4 IPv6 网络地址规划 ……112
5.4.1 设计思路 ……112
5.4.2 设计原则 ……116
5.4.3 设计要素 ……116
5.5 网络安全及业务溯源 ……121
5.5.1 IPv6 网络安全 ……121

5.5.2 IPv6 业务溯源……124
5.5.3 运营级协议转换技术网络安全……124
5.5.4 运营级协议转换技术业务溯源……127
5.6 总体演进策略建议……129

第 6 章 内容和应用过渡技术部署……132
6.1 IDC 的双栈升级改造方案……132
6.2 典型内容和应用迁移方案分析…133
6.2.1 网页浏览类应用迁移方案……133
6.2.2 FTP 应用迁移……140
6.2.3 电子邮件应用迁移……141
6.2.4 即时通信应用迁移……145
6.2.5 流媒体应用迁移……147
6.3 IPv6 与物联网应用……150
6.3.1 物联网概念与架构……150
6.3.2 物联业务驱动的 IPv6 演进……153
6.4 内容和应用迁移策略建议……157
6.4.1 大中型 ICP 的升级双栈改造……157
6.4.2 小型 ICP 双栈的升级改造……157

第 7 章 终端演进……158
7.1 当前各种终端支持 IPv6 能力现状……158
7.2 固网终端……158
7.2.1 IPv6 对固网终端的技术要求……159
7.2.2 IPv4/IPv6 过渡技术的演进策略……161
7.3 PC 操作系统和客户端……164
7.4 手机终端……165
7.5 3G 上网卡……166
7.6 IPTV 机顶盒……166
7.7 物联网终端……167
7.7.1 物联网终端的分类……167
7.7.2 物联网网关……168
7.7.3 短距离无线通信技术……170
7.8 终端演进策略建议……176
7.8.1 固网终端演进策略……176
7.8.2 移动终端演进策略……177

第 8 章 IPv6 过渡技术部署实例……178
8.1 分布式 NAT444……178
8.1.1 实验内容与目标……178
8.1.2 网络结构……178
8.1.3 业务流程……179
8.1.4 地址分配……180
8.1.5 NAT 及其穿越……180
8.1.6 用户溯源……181
8.1.7 设备具体配置……181
8.2 集中式 NAT444……184
8.2.1 实验内容与目标……184
8.2.2 网络结构……185
8.2.3 业务流程……186
8.2.4 地址分配……186
8.2.5 NAT 及其穿越……186
8.2.6 用户溯源……186
8.2.7 设备配置……186
8.3 分布式 DS-Lite……186
8.3.1 实验内容与目标……186
8.3.2 网络结构……187
8.3.3 业务流程……188
8.3.4 地址分配……188
8.3.5 NAT 及其穿越……189
8.3.6 用户溯源……189
8.3.7 DNS 解析……189
8.3.8 设备具体配置……189
8.4 集中式 DS-Lite……191

8.4.1 实验内容与目标 ····· 191
8.4.2 网络结构 ····· 191
8.4.3 其他 ····· 192
8.5 IPv6 过渡技术实验小结 ····· 192

第 9 章 IPTV 业务 IPv6 过渡部署实例 ····· 193
9.1 实验概况 ····· 193
9.1.1 IPTV 现有业务系统介绍 ····· 193
9.1.2 实验目标 ····· 194
9.2 过渡部署方案及流程 ····· 194
9.2.1 现有系统改造方案 ····· 194
9.2.2 改造后的点播业务流程 ····· 196
9.2.3 改造后的直播业务流程 ····· 197
9.2.4 改造后的增值业务流程 ····· 198
9.2.5 改造后的上网业务流程 ····· 199

缩略语 ····· 201

第 1 章
IPv6 简介

1.1 IPv4 向 IPv6 过渡的必然性

IP 地址即互联网地址，是用来标志互联网终端的逻辑地址，具有唯一性。我们现在正在应用的网络互联协议是瑟夫 20 世纪 70 年代创立的一种名为 IPv4（Internet Protocol version 4）的 32 位地址，也是世界上被广泛使用的网络互联协议的最重要组成部分。该机制通过编码的方式用 32 位二进制数来区分识别不同的主机，理论上最多有 $2^{32}-1$ 个地址，最多可以容纳 $2^{32}-1$ 台主机（不包括虚拟子网）连入互联网，总地址数为 42 亿多。在当时，谁也没有想到这么大的容量也会用完。

早在 18 年前，有互联网专家就指出，由于 IPv4 技术本身存在着网络地址资源浪费的问题，加上网络的迅猛发展，IP 地址将很快耗尽。专家指出，在采用 A、B、C 3 类编址方式后，虽然方便了 IP 地址的管理，但浪费了上千万的地址，特别是 B 类地址。例如，一个需要 300 部电脑连上互联网的公司，使用一个 C 类地址（254 个地址）是不够的；使用两个 C 类地址会使路由选择表爆涨，从而影响了外部路由协议的性能；另外的选择是使用一个 B 类地址（65534 个地址），但却浪费了 65234 个地址。

而且，美国一些大学和公司占用了大量的 IP 地址。国际电信联盟副秘书长赵厚麟日前表示，美国人口只有 3 亿，但掌握的 IP 地址接近全球总数的 40%，其中大部分一直被闲置。

另一方面，在互联网快速发展的国家，如欧洲很多国家，以及日本和中国由于得不到足够的 IP 地址，最后导致互联网地址耗尽和路由表爆涨。到目前为止，A 类和 B 类地址已经用完，只有 C 类地址还有余量。

另外，目前占有互联网地址的主要设备早已由 20 年前的大型机变为 PC，宽带业务正在飞速发展中，访问 Internet 的终端类型越来越丰富，除了传统的 PC，越来越多的其他设备也要连接到互联网上，包括 PDA、手机、汽车、各种家用电器等。特别是手机，为了向第三代移动通信标准靠拢，几乎所有的手机厂商都在向国际因特网地址管理机构 IANA 申请地址。IANA 的 IPv4 地址在 2011 年 2 月 3 日已经分配完毕，5 个 RIR 管理的地址空间中，RIR 亚太地区已经分配完毕。而随着物联网的发展，将会使所有的信息设备智能化联网，如传感器、RFID 读卡器、智能楼宇设施等要接入互联网，都需要有 IP 地址的支持，物联网的发展将需要更多的地址空间，IPv4 显然已经无法满足这些要求。

为解决 IPv4 地址枯竭的现象，IPv6（Internet Protocol version 6）应运而生。

IPv6 是 IETF 设计的用于替代 IPv4 的下一代互联网协议。其最主要的特点就是地址空间几乎无限。与 IPv4 不同的是，该协议规定的地址长度为 128bit，地址空间比 IPv4 增大了 2^{96} 倍，可以很好地解决 IP 地址分配不足的问题。目前 IPv4 正在向 IPv6 过渡。

1.1.1 国际国内 IPv4 地址枯竭形势

2011 年 2 月 3 日，国际互联网名称和编号分配公司（ICANN）公布了一条“可以载入史册的”新闻：“最后一批 IPv4 地址今天分配完毕，IP 地址总库已经枯竭。”该消息一经传出，立即引起全球网民的高度关注。

谷歌副总裁兼首席互联网顾问文顿•瑟夫有“互联网之父”之称，他曾预言“IP 地址即将在几周内用完”，没想到在他发表此言论的两周后，这句话就得到了印证。

我们从中国互联网络信息中心（CNNIC）了解到，截至 2010 年年底，我国持有的 IPv4 地址数量为 2.77 亿，还不到全球 IP 地址总数的 10%，而中国互联网用户达 4 亿多，远远超过了中国 IP 地址的容量。如今，IP 地址耗尽，中国网民会不会面临无网可上的困境？

CNNIC 的 IP 地址业务负责人赵巍在 2011 年年初曾经认为：“IP 地址用完的影响不会马上显现。”在我国拥有的 IPv4 地址中，运营商手中掌握了部分 IPv4 地址可以供使用，但随着运营商业务的发展，如果不采取措施，运营商手中的 IPv4 地址仅仅能够支撑 1～2 年。

专家预计，IP 地址资源的紧缺与各国的加速争抢会形成恶性循环，未来企业获得 IP 地址的成本会越来越高。DCCI（中国互联网数据中心）总经理胡延平曾对媒体表示：“IPv4 地址枯竭对于提供互联网数据（IDC）服务的企业会造成困扰。”比如某企业要在 IDC 上托管服务器，IDC 会分配一个 IP 地址。现在 IPv4 地址枯竭了，企业可能就要共享一个 IP 地址，IDC 的服务就会大打折扣。绝大多数 IDC 厂商由于其 IP 资源需要通过电信运营商购买，价格已经从原来的几十元升至几百元，到达最终用户的手中的价格自然更贵。

IPv4 地址短缺虽然短期内影响不大，不过专家提醒，如果长期 IP 地址耗尽的问题得不到解决，对于普通网民来说，IP 地址资源的紧缺可能导致上网费用上升。对于新入网用户来说，如果分不到 IPv4 地址，那么访问现有的互联网资源和服务将受到很大限制；对于运营商和内容提供商来说，没有足够的 IPv4 地址会制约其提供新的服务。

1.1.2 应对 IPv4 地址枯竭的手段

为了应对地址枯竭的严峻形势，人们早就开始研究应对的手段，概括来说主要有存量地址的挖潜、私网地址的引入、引入 IPv6 地址 3 种。

（1）存量地址的挖潜

对存量地址的挖潜就是利用管理手段和技术手段重新梳理已经使用的 IPv4 地址，通过回收、合并、调剂、复用等方法使 IPv4 地址的分配更加合理，使用的效率更高。原有的地址规划一般是通过需求预测来分配的，但后来的使用情况与事先的预测不完全一致，多少不均。因此需要运用管理手段来重新进行地址规划。但这种方式在一个运营商或企业之内可以根据情况采用，不同运营商或企业之间很难实行。目前运营商内部的地址挖潜主要是：零散地址回收、合并，控制专线互联网业务每个用户地址的使用数量，推动萎缩型业务退网，同类型业务地址之间的调剂、复用，等等。

通过技术手段进行存量地址的挖潜主要是通过动态地址分配和无域间路由（CIDR）来实现。很多业务都采用了动态地址分配的技术，即在用户需要公网地址的时候，从地址池中临时分配一个地址，用户下线时释放该地址。采用这种方法，公网地址数量的配置需要根据用户上网行为合理配置，以达到提高地址利用率的目的。另外一种方法，无域间路由技术则已经为大家熟知，并广泛运用于地址分配领域。它支持任意长度的子网掩码，可以按需一级一级地分配地址。

（2）私网地址的引入

私网地址是应对地址枯竭的一种非常快速和有效的手段。私网地址在 IPv4 地址中定义了 10/172/192 3 段，不同网络的私网地址可以重复，因此不能在公网上路由。私有地址用户访问公网时必须使用 NAT 技术转换为公网地址。NAT 技术可以将一个地址加上端口号后分给多个用户使用，大大节约了 IPv4 地址的使用。根据现网统计，1 个公网地址可转换 60 个私网地址，即满足 60 个在线宽带用户的地址需求。这种方法大大节约了公网 IPv4 地址，解决了私网地址之间冲突的问题，并且隐藏了私网的拓扑结构，相对安全和易管控。但是 NAT 也带来了一系列的问题：增加了报文的延迟；增加了维护和配置的复杂性；破坏了互联网端到端的特性，导致某些应用无法使用（如 FTP、即时通信等）；用户溯源困难；大容量的 NAT 设备易成为网络瓶颈，等等。

（3）引入 IPv6 地址

IPv6 是针对 IPv4 地址的不足而设计出来的一套协议，对网络架构、应用等基本无影响。IPv6 地址有 128 位，2^{128} 个地址的数量远远大于 32 位的 IPv4 地址数。海量的地址空间不仅可以支持人之间、人物间的通信，也可以支持物物间的通信，极大地支持了物联网业务的发展。它在安全性和扩展性上也较 IPv4 地址做了改进，但是它在协议标准、设备及终端支持、产业链等方面尚未完善。

1.1.3　产业链各方向下一代互联网演进策略

2011 年，IPv6 产业链整体处在规模试点阶段。从总体发展来看，呈“纺锤形”不均衡发展，即：网络准备度好，应用匮乏，终端改造成本高。

（1）固网终端

终端改造有难度，改造成本高。主流的 PC 操作系统 Windows XP 不支持 IPoEv6/PPPoEv6，需开发软终端才能完成 IPv6 认证和接入，Windows 7、Windows Vista 支持 IPv6，但非主流；桥接型网关可透传双栈业务，不需要改造，现网路由型家庭网关（HG）不支持 IPv6，路由型家庭网关改造成本高。

（2）移动终端

业界智能手机主流操作系统（如 Symbian、Android、iOS）均支持 IPv6 协议栈，但手机能否实现 IPv6 接入，还取决于手机芯片和软件平台是否适配 IPv6 操作系统。目前只有个别手机支持 IPv6，C 网和 TD 手机很少支持 IPv6。数据卡桥接模式支持 IPv6 接入，WCDMA、cdma2000、TD-SCDMA 及 LTE 的主流芯片厂商高通、威盛、联芯等也有计划后续支持 IPv6。

（3）承载网络

网络双栈改造支持度高。经过中国电信、中国移动、中国联通两年多的测试和现网试点验证，现网 70%左右的路由器和 BRAS 支持双栈升级，城域核心、骨干网双栈改造技术成熟，

接入网设备类型复杂，IPv6 试点初期可二层透传双栈业务，后期 IPv6 双栈改造成本高。核心网主流 GGSN/PDSN 设备支持双栈改造，RAN 只需 IPv4 透传 IPv6 即可，支撑网平台支持双栈改造。

（4）业务和应用

当前，IPv6 的应用整体较少，整体占比小于 0.5%，业务和应用匮乏，阻碍 IPv6 产业链均衡发展。与 3G、4G 不同，IPv6 只是网络层基础协议，目前还没有带来巨大商业价值的具体业务，ICP 和运营商向 IPv6 迁移还存在成本和投入风险，导致 ICP 对发展 IPv6 不积极，运营商对 ICP 开通 IPv6 业务影响有限。国内当前开通 IPv6 服务的网站主要集中在教育、科研和校园网，未见主流 ICP 开通 IPv6 的明确计划。

1.2 IPv6 的发展历程及现状

作为 IPv4 的“接班人”，IPv6 地址由原来的 32 位扩充到 128 位。也就是说，IPv6 理论上可以产生近乎无数个 IP 地址，确切地说，是 34 后面带 36 个零。这个天文数字意味着，如果全球有 1800 亿亿网民，每人拥有 1800 亿亿互联网终端，所有这些终端同时上网，IP 地址仍有富余，可以“让地球上的每一粒沙子都具有一个 IP 地址”。

但是从 1995 年准备到现在，IPv6 推广情况并不理想。虽然随着全球 IPv4 地址告急，越来越多的国家启动 IPv6 发展计划，欧、美、韩、日、澳及人口大国（如中国、巴西、俄罗斯）发展最快，均有或正在实施 IPv6 实验网，但全球当前 IPv6 用户渗透率仍低于 1%。

目前，全球仅有几家运营商启动 IPv6 试商用，主要是 NTT、Verizon 和 AT&T，西欧和中国处于第二梯队，IPv6 已经开始从实验网走向现网试点，今年还有可能实现小规模试商用。

我国于 2003 年在国家发改委的牵头下，启动 CNGI 实验网建设。中国移动、中国电信、中国联通等运营商已经在国内逐步部署了具有一定规模的 IPv6 骨干网络。其作为国家级 IPv6 实验网骨干，已经实现与国际 IPv6 互联网对接。目前，下一代互联网已列入我国“十二五”发展纲要中，工业和信息化部在向运营商征询“十二五”互联网的规划意见，IPv6 是下一代互联网规划的重点。

在 IPv6 标准的制定上，通过各标准化组织的协作和联合，IPv6 协议标准已逐步完善，过渡技术还需要探索验证。

与 IPv6 相关的标准化组织包括 IETF（因特网工程任务组）、ITU-T（国际电信联盟标准化部门）、3GPP（第三代合作伙伴计划）等，各个组织有不同的着眼点和侧重点。IETF 主要致力于 IPv6 在因特网中的应用，ITU-T 主要致力于 IPv6 在电信网中的应用，3GPP 主要致力于 IPv6 在核心网中的应用。作为下一代网络中的核心协议，IPv6 标准的制定需要多个相关组织的协作与联合。

IPv6 标准体系主要包括资源（编址和域名等）、网络、应用、安全和过渡技术几方面。

① 资源类和网络类协议标准主要由 IETF 负责。核心标准已完善，可以满足 IPv6 组网需要。

② 应用类标准由 IETF、3GPP、ITU 等多个组织负责。移动宽带和固定宽带应用基本完善，但 NGN 类没有完成。

③ 安全类标准还不完善，特别是网络可信度方面。

④ 过渡技术类标准中 IPv6 迁移的过渡技术标准，如双栈和隧道已完善，但完善的电信级网络地址转换（NAT）技术标准尚未完成。

1.3　IPv6 技术的新特点

制定 IPv6 的专家们充分总结了早期制定 IPv4 的经验以及互联网的发展和市场需求，认为下一代互联网协议应侧重于网络的容量和网络的性能。IPv6 继承了 IPv4 的优点，摒弃了它的缺点。IPv6 与 IPv4 是不兼容的，但它同所有其他的 TCP/IP 协议簇中的协议兼容，即 IPv6 完全可以取代 IPv4。同 IPv4 相比较，IPv6 在地址容量、安全性、网络管理、移动性以及服务质量等方面有明显的改进，是下一代互联网可采用的比较合理的协议。IPv6 协议的技术优势主要体现在以下几方面。

1.3.1　地址管理

（1）地址空间扩大

极大的地址空间是 IPv6 最令人瞩目的特征。IPv6 采用 128bit 地址长度，地球表面平均每平方米可以分配到超过 1mol（6.02×10^{23}）地址，即使 H（地址分配效率系数，H =log 地址数/位数）保持在 0.22～0.26 的水平，也足够在可以预见的未来几十年内使用。地址长度的增加，并不仅仅是地址数量的增加，还给地址分配、自动配置等方面带来质的提高。

（2）地址分配合理

IPv6 不必受地址资源限制而捉襟见肘，所以在地址格式设计、地址分配等方面尽可能做到合理，同时为不可预见的问题留有足够的余地。

IPv6 地址分为单播、多播、任播 3 类，并为将来保留了 85%的地址。根据 IPv6 单播地址作用范围的不同，又分为可聚合全球单播地址、链路本地地址和站点本地地址。此外，还有一些特殊地址也属于单播地址，如内嵌 IPv4 的 IPv6 单播地址，为 NSAP 和 IPX 保留的地址等。

IPv6 取消了 IPv4 地址类的概念，64 位作为网络号，64 位作为主机号。网络号根据需要进一步划分，对地址格式做了详细定义，方便了聚类和路由，预留了一定地址位数以应付未能预见的问题。IPv6 地址为 ISP 所有，避免了 IPv4 地址用户所有带来的地址浪费和 ISP 路由聚类困难。

（3）自动配置和移动支持

自动配置是 IPv6 的重要进步，分为无状态自动配置和有状态自动配置两种方式。一个节点只需要将自己链路层的 IEEE EUI-64 地址作为主机号，结合 ISP 提供的本地网络号，就能够通过 IPv6 的自动配置得到唯一的 IPv6 地址，实现“即插即用”。

“即插即用”简化了网络管理和控制，使得移动 IPv6 比 IPv4 更容易实现和管理，同时移动 IPv6 还对代理联络和安全策略做了大量改进，邻居发现等新技术也给移动 IPv6 以强有力的支持。

1.3.2　改进报头

IPv6 是对 IPv4 彻底改革而不是修补的重要体现，是对数据报报头的改进，这也是 IPv6 在其他方面重大改进的基础。IPv4 报头不定长且结构复杂，主机和路由器都难以提高处理效

率。IPv6 简化了基本报头，降低了处理复杂度，并使用扩展报头提高适应性和扩展性。

（1）简化报头

IPv6 省去了 IPv4 报头中的部分字段以简化结构，IPv6 地址长度是 IPv4 的 4 倍，但报头长度只有 IPv4 的 2 倍。IPv6 的基本头部为固定长（40Bytes），无需头标长度标识。IPv6 只支持端点对端点分段，不再需要标识符、标志和偏移量字段。IPv6 取消了头标校验和字段，以简化对数据报报头的处理。

IPv6 对 3 个字段重新命名，并赋予新的含义。IPv4 数据报总长度由 IPv6 的有效载荷长度代替。IPv6 把协议类型字段重新命名为“下一报头”，指明在 IPv6 基本报头后面报头的类型，它可能是一个扩展报头或数据净荷。IPv6 中用跳数限制表示取代了 IPv4 的生存期概念。

（2）扩展报头

IPv6 的扩展报头模式借鉴了 IPv4 任选项，将报头中不是每个节点都用到的字段改为可选项处理，附加在基本报头后构成扩展报头。大多数 IPv6 扩展报头不受路由器的检查，从而提高了路由器的转发效率。目前 IPv6 已定义的可选扩展报头有：逐跳可选报头、接收端可选报头、选路可选报头、分段可选报头、验证可选报头、封装安全净载荷可选报头和上层协议可选报头。一个 IPv6 数据报可以根据需要携带零个、一个或多个扩展报头，提供了最大的灵活性。IPv6 扩展报头的结构类似于数据结构中的指针链表，基本报头和每一个扩展报头都包含下一个报头（Next Header）字段，每一个扩展报头都由特定的下一个报头值（Next Header Value）来确定。

1.3.3 改进路由

（1）ICMPv6

ICMP 并不是 IP 层路由功能的一部分，但 IPv6 很多路由方面的新特征都依赖于 ICMP。IPv6 对 ICMP 做了大量改进，升级为 ICMPv6。ICMPv6 具备目前 ICMP 的基本功能，并综合了原 IPv4 中分属不同协议完成的功能。多播收听者发现（MLD）用 ICMPv6 消息取代了 IPv4 所用的因特网组管理协议（IGMP），管理多播成员资格，使得效率和安全性有了明显提高。

ICMPv6 实现的更重要的新功能是邻居发现协议（NDP）。NDP 是 IPv6 协议的一个基本的组成部分，用来管理同一链路上节点间的通信。NDP 取代了数据链路层的 ARP，抑制了广播风暴，提高了安全性。NDP 能够完成邻居发现和路径 MTU，为 IPv6 的源主机分段提供信息。路由器通过 NDP 宣告邻接路由器转发数据报，通知发送端重定向，实现最佳路由。此外，NDP 还为自动配置提供网络前缀等参数，检测地址可达性和重复地址。

（2）从 BGP-4 到 IDRP

IPv6 域间路由最大的改进在于 IDRP 替代了 BGP-4。由于 BGP（Border Gateway Protocol，边界网关协议）对 32 位的 IPv4 优化程度相当高，很难为 IPv6 升级，因此 IPv6 所使用的外部网关协议以 IDRP（Inter Domain Routing Protocol，域间路由选择协议）为基础。

IDRP 和 BGP-4 的主要区别有：

① BGP 报文通过 TCP 交换，IDRP 协议单元直接通过数据报服务来传递；

② BGP 是一个单地址族协议，IDRP 可以使用多种类型的地址；

③ BGP 使用 16 位的自治系统编号，IDRP 使用变长的前缀来标识一个域。BGP 描述的

是路径所通过的自治系统编号的完整列表，而 IDRP 能对这个信息进行聚集。

（3）源主机分段

实践证明 IPv4 逐跳分段并不理想，增加了路由器负担，一个分段的丢失会导致所有分段重传。IPv6 分段只发生在源节点，简化了报头并减少了路由器的分段开销。IPv6 要求各节点间 MTU 的最小值为 1280Bytes，兼顾了网络效率和旧设备投资，并要求所有节点支持路径 MTU 发现，根据链路状况选择最佳分段大小。

通过以上路由方面的改进，再加上地址格式的变化和报头的简化，大大降低了主机和路由器的复杂性和负荷。据 Cisco 资料表明，在 Cisco 2600 系列路由器中配置的 IPv4 内核为 2.17MB，如计算存放路由表的工作区则升至 3.2MB，而配置 IPv6 的内核时，其内核仅为 1.69MB，加上工作区也不过 2.7MB，而且路由效率有明显提高。

1.3.4　安全机制

（1）IPSec

IP 层最初没有安全性方面的考虑，认证和保密都是由上层协议来完成的，很多黑客攻击就是利用了 IP 层这方面的不足。IPSec 协议族就是为在 IP 层实现安全性而设计的，IPv4 和 IPv6 都可以使用，所不同的是 IPSec 必须加以修改或改进才能应用在 IPv4 中，而 IPSec 是 IPv6 的重要组成部分，IPv6 所有应用从一开始就具有这些安全特性。

（2）AH 和 ESP

安全性一般认为有 3 个要求：身份认证、保密性和完整性，IPSec 的目标是实现前两个要求。RFC 1826 定义了认证头部（AH，Authentication Header），RFC 1827 定义了加密安全净载荷（ESP，Encrypted Security Payload）。前者提供认证机制，通过认证过程保证接收者得到的数据报来源是可靠的，而且在传输过程中没有被偷换。后者使用密钥技术，保证只有合法的接收者才能读取数据报的内容。IPv6 使用扩展报头实现 AH 和 ESP。

（3）安全关联

认证和加密要求发送者和接收者就密钥、认证和加密算法以及一些附属细节达成一致。这套约定机制组成了发送者和接收者之间的一种安全关联。接收者在接收到数据报后，只有在能将其与一种安全关联的内容相关联时，才能对其进行验证和解密。所有的 IPv6 验证和加密数据报都带有安全参数索引（SPI，Security Parameter Index）。

1.3.5　增强服务质量

IPv4 对不同的用户和不同的应用都不加区分，采取尽力而为的传输方式，服务质量（QoS，Quality of Service）难以保证。而 IPv6 通过设置优先级、数据流标签等方式，对 QoS 特别是 VoIP 等实时数据流的传输提供了很好的支持。

（1）优先级

IPv6 报头的通信流类型（Traffic Class）可理解为优先级标识，数值越大优先级越高。

IPv6 将业务分为两大类，阻塞控制业务和非阻塞控制业务，前者在网络阻塞时流量会降低，而后者不会变化，用于声音和视频传输。IPv6 为阻塞控制业务分配编号 0～7 的优先级，推荐电子邮件为 2，大量数据传输（如 FTP）为 4，交互式（如 Telnet）为 6。对于非阻塞控制业务，根据媒体质量的不同分配编号 8～15 的优先级。

（2）流和资源预留

IPv6 增加了数据流标签，发送端将需要路由器特殊处理的数据报在数据流标签字段加以标识。数据流标签是 20 位的随机数，同一节点同一数据流所产生的数据报具有相同的流标签。数据流标签还可以和路由选择扩展报头同时使用。为了更好地实现流传输，还开发了资源预留协议（RSVP）。RSVP 是一个接收者驱动（receiver-driven）协议。接收者通过发送 RSVP 报文选择接收源以及准备预留的带宽和费用等，达到最理想的效费比。

第 2 章
IPv6 技术基础

2.1 IPv6 地址介绍

IPv6 具有 128 位的地址长度，是 IPv4 地址长度的 4 倍。IPv6 所拥有的地址数量更是 IPv4 地址的 2^{96} 倍。IPv6 超乎想象的地址数量能很好地解决地址缺乏的难题。本节将介绍 IPv6 地址的结构、寻址模型以及地址分类。

2.1.1 IPv6 地址结构

IPv6 地址有 3 种格式：首选格式、压缩格式和内嵌 IPv4 地址的 IPv6 地址表示。

（1）首选格式

IPv6 的 128 位地址每 16 位划分为 1 段，每段被转换为一个 4 位十六进制数，并用冒号（:）隔开，即冒号十六进制法。

假设 128 位 IPv6 地址为：

0010000111011010000000001101001100000000000

0000000010111100111011000000010101010100000

00001111111111111111000101000100111000101101O，

则冒号十六进制表示法为：21DA:00D3:0000:2F3B:02AA:00FF:FE28:9C5A。

（2）压缩格式

冒号十六进制法表示的 IPv6 地址中，当存在一个或多个连续的段为 0 字符时，可以用一对冒号（::）来表示，但一个 IP 地址中只能出现一个（::）；当一段中高位为 0 时，可以略去高位。

假设一个 IPv6 地址为：2F02:0030:0000:0000:0001:0000:0000:0002，则应用压缩格式可以表示为：2F02:30::1:0:0:2。

（3）内嵌 IPv4 地址的 IPv6 地址

这是 IPv4 向 IPv6 过渡机制中使用的方法，在这种表示方法中，前 96 位用冒号十六进制表示，后 32 位用点分十进制表示。内嵌 IPv4 地址的 IPv6 地址格式有两种类型。

① IPv4 兼容地址：用于在运行 IPv4 和 IPv6 两种协议的节点使用 IPv6 进行通信，节点在 IPv4 路由结构之上构建动态 IPv6 隧道。表示方法为：::w.x.y.z，其中 w.x.y.z 是 IPv4 地址。

该地址形式已经在 RFC 4291 中被废弃。

② IPv4 映射地址：用于将仅支持 IPv6 访问而不支持 IPv6 协议的节点地址表示成 IPv6 地址，表示方法为：::FFFF:w.x.y.z。该地址在 6PE、6vPE 中有应用。

（4）地址前缀

IPv6 的地址前缀仍然用文本表示法，一个 IPv6 地址前缀用如下形式表示：IPv6 地址/前缀长度。

这里前缀长度是一个十进制的数，表明地址最左端连续的多少比特构成该地址的前缀。假设节点 IPv6 地址为 1F02:30::1:0:0:2，则节点子网号为 1F02:30:0::/40，可以简写为 1F02:30::1:0:0:2/40。

2.1.2 IPv6 寻址模型

IPv6 地址被分成两个部分：子网前缀和接口标志符。每个单播地址标识一个单独的网络接口。IPv6 地址被指定给网络接口而不是节点，因此一个拥有多个网络接口的节点可以具备多个 IPv6 地址，其中任何一个 IPv6 地址都可以代表该节点。尽管一个网络接口可以与多个单播地址相关联，但一个单播地址只能与一个网络接口相关联，一个网络接口必须至少具备一个单播地址。

需要注意的是：如果点到点链路上的任何一个端点都不需要从非邻居节点接收和发送数据，则不需要 IPv6 地址；如果硬件有能力在多个网络接口上正确地共享其网络负载，则这些网络接口可以共享 IPv6 地址。

2.1.3 IPv6 地址分类

IPv6 地址有 3 种类型：单播（unicast），多播（multicast）和任播（anycast）。

① 单播地址：一个单接口的标识符，送往一个单播地址的分组被送至该地址标识的一个接口上。

② 多播地址：一组接口的标识符，送往一个多播地址的分组被传送至该地址标识的所有接口。

③ 任播地址：一组接口的标识符，送往一个任意播地址的分组被传送至该地址标识的接口之一（“路由最近”）。

注意：IPv6 不带有广播地址，解决办法为使用所有节点（ALL-NODES）的多播地址。

1. 单播地址

IPv6 单播地址根据作用范围的不同主要分为：链路本地地址（Link-local Address），站点本地地址（Site-local Address），可聚合全球单播地址（Aggregatable Global Unicast Address）。此外，还有一些特殊用途的单播地址，如唯一本地地址（Unique Local Address）、未指定地址、本地回送地址（Loopback Address），以及内嵌 IPv4 地址的 IPv6 单播地址等。

（1）链路本地地址

链路本地地址用在同一本地链路的节点之间，带有链路本地源或目的地址的数据分组不转发到其他链路上，主要用于局域网管理。在 IPv6 的邻居节点之间的通信协议（如邻居发现协议）中广泛使用此地址。

链路本地地址的地址格式：前 10bit 为 0xFE80，中间 54bit 为“0”，最后 64bit 为接口 ID。其地址格式如图 2-1 所示。

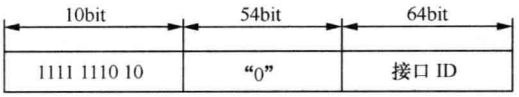


图 2-1　链路本地地址格式

获取 IPv6 接口 ID 最常用的方法是生成修正的 IEEE EUI-64 接口 ID，这种生成方式因链路类型而异，例如基于以太网接口和光纤分布式数据接口的生成方式是不同的，这里只介绍以太网接口的基于 MAC 地址生成方式。

基于 MAC 地址的修正 EUI-64 接口 ID 生成方式如下。

MAC 地址共 48 位，在前 24 位和后 24 位之间添加 16 位的 0xFFFE。

MAC 地址第 7 位（从高位起）反转取“1”。

图 2-2 给出了其转换过程，因为物理网卡的 MAC 地址是全球唯一的，所以使用这种方式生成的接口 ID 也是全球唯一的。

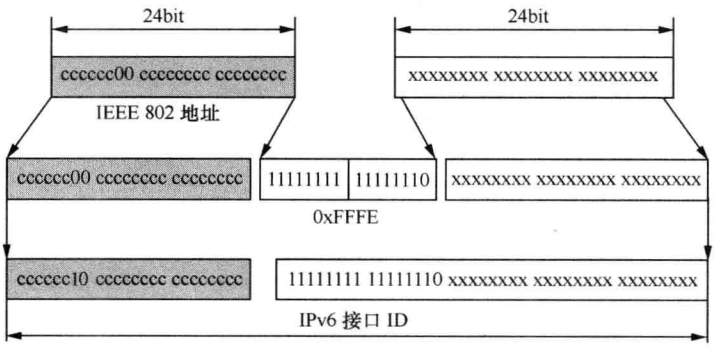


图 2-2　EUI-64 地址转换

（2）站点本地地址

站点本地地址用在单一站点内部，带有站点本地源或目的地址的数据分组不转发到其他站点上。

站点本地地址的地址格式：前 10bit 为 0xFEC0，中间 54bit 为子网 ID，由站点管理员分配，最后 64bit 为接口 ID。其地址格式如图 2-3 所示。

目前站点本地地址已经被废弃，见 RFC 3879。

（3）可聚合全球单播地址

带有可聚合全球单播地址的数据分组可被传送到全球 IPv6 网络的任何部分，相当于 IPv4 的公网地址。RFC 3513 给出了新的 IPv6 可聚合全球单播地址通用格式，如图 2-4 所示。

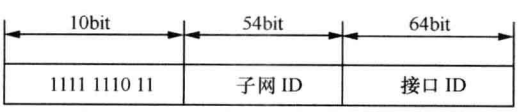


图 2-3　站点本地地址格式

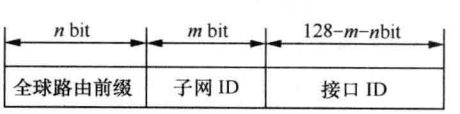


图 2-4　可聚合全球单播地址格式

全球路由前缀是分配给站点的一个典型层次结构值，子网 ID 是一个站点内子网的标识。全球路由前缀由 RIR 和 ISP 设计为分层结构，子网域由站点管理员设计为分层结构。

由于 RFC 3513 要求所有的单播地址有 64 位接口 ID 并具有改进 EUI-64 格式的结构，即要求 $n+m=64$，为进一步明确 IPv6 全球单播地址格式，RFC 3587 给出了新的格式，如图 2-5 所示。

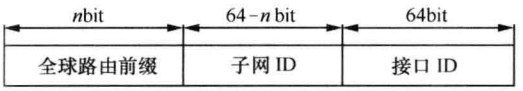


图 2-5　可聚合全球单播地址新格式

RFC 3587 规定只有 001 前缀格式供 IANA 分配，其余的全球单播地址空间（大约是 IPv6 地址空间的 85%）要保留下来，以备将来定义和使用，暂时不再由 IANA 分配。所以，由 IANA 代理的 2000::/3 前缀的全球单播地址就成了如图 2-6 所示的格式。

001	全球路由前缀	子网 ID	接口 ID
运营商前缀		城域网前缀	接口标识符
网络部分			节点部分

图 2-6　供分配的可聚合全球单播地址格式

（4）唯一本地地址

唯一本地地址在全球 Internet 中不可路由，仅在站点内部可路由，用于站点内部通信，它有如下特性。

① 固定的地址前缀格式，便于站点边缘过滤。

② 全球唯一（唯一性概率很高）。

③ 站点合并无需对接口地址重编号。

④ 独立于 ISP，任何人都可以随意使用，站点内部通信无需 Internet 连接。

⑤ 路由协议与应用协议无需改变，应用协议认为其是全球单播地址。

唯一本地 IPv6 地址格式如图 2-7 所示。

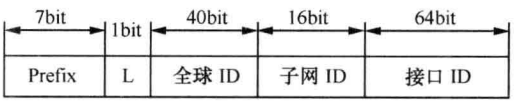


图 2-7　唯一本地 IPv6 地址格式

其中各字段的含义如下。

① Prefix：前缀格式固定为 FC00::/7，该前缀指明地址为唯一本地 IPv6 地址。

② L：指明全球 ID 的产生方式，置“1”表明本地生成全球 ID；置“0”未定义，留作将来用。

③ 全球 ID：40bit，目前由伪随机算法生成，全球唯一性概率非常高。

④ 子网 ID：16bit，在划分子网时使用。

⑤ 接口 ID：64bit，通常由 EUI-64 生成。

唯一本地地址在 RFC 4193 中定义，其在构建 IPv6 VPN 的应用中前景光明。唯一本地地址避免了 IPv4 中私有地址泄漏的问题，而一旦地址泄漏也由于其唯一性不会产生地址冲突。

（5）特殊地址

分为未指定地址和本地回送地址。

未指定地址即为全“0”地址，表示一个地址的缺失。全“0”地址不能用作目的地址，仅当初始化主机不知其地址时，其发送的 IPv6 分组的信源地址即为全“0”地址。

本地回送地址为::1，不能分配给任何物理接口，本地回送地址就相当于接口本身。当接口向本地回送地址发送数据分组时，数据分组不能发送出节点，只有该接口自己能收到，通常用于节点自测。

（6）内嵌 IPv4 地址的 IPv6 单播地址

在上文 IPv6 地址格式中已有介绍。

2．多播地址

一个多播地址是一组接口的标识符，一个节点可能属于任意数量的多播群。多播地址不能用在 IPv6 数据分组的源地址域中，也不能出现在任何路由头中。

多播地址格式和各字段说明如图 2-8 所示。

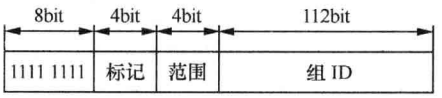


图 2-8　IPv6 多播地址格式

其中各字段的含义如下。

① 0xFF：最高 8bit 为 1111 1111，标识此地址为多播地址。

② 标记：4bit，从高到低各字段依次为保留位、R 位、P 位以及 T 位，各位的取值如下。

- 最高位保留位必须置“0”。
- R 位取“0”表示非内嵌 RP 的多播地址；取“1”则表示内嵌 RP 的多播地址，此时 P、T 位也必须置“1”。
- P 位取“0”表示非基于单播前缀的多播地址；取“1”则表示基于单播前缀的多播地址，此时 T 位也必须置“1”。
- T 位取“0”表示永久分配的多播地址；取“1”则表示非永久分配的多播地址。

③ 范围：4bit，用来标识此多播组的应用范围，其取值及含义如下。

- 0、3、F：保留（reserved）。
- 1：接口本地范围（interface-local scope）。
- 链路本地范围（link-local scope）。
- 管理本地范围（admin-local scope）。
- 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）。

④ 组 ID：112bit，多播组标识号，用来在由 Scope 字段所指定的范围内唯一标识多播组。

目前已永久分配了一些常用 IPv6 多播地址，如“所有节点多播地址”FF02::1（链路本地）和“所有路由器多播地址”FF02::2（链路本地）等。一些比较特殊的多播地址在 IPv6

中也有应用，我们介绍被请求节点多播地址（Solicited-node Multicast Address）、基于单播前缀的 IPv6 多播地址以及内嵌 RP 地址的 IPv6 多播地址。

（1）被请求节点多播地址

被请求节点多播地址主要用于重复地址检测和获取邻居节点的链路地址。

被请求节点多播地址生成算法：选取一个地址（单播或任播）的低 24 位，将这 24 位附加到前缀 FF02::1:FF00/104 之后。

被请求节点多播地址受限于本地链路范围。一个节点对于其分配的每个单播和任播地址都自动启用一个与之对应的被请求节点多播地址。

（2）基于单播前缀的 IPv6 多播地址

基于单播前缀的 IPv6 多播地址中包含了其多播源网络的单播地址前缀，适用于 SSM 多播模型，无需大量分配协议，通过基于单播前缀的 IPv6 多播地址网即可动态获得多播地址。

基于单播前缀的 IPv6 多播地址格式如图 2-9 所示。

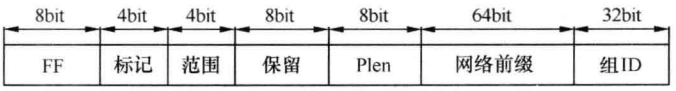


图 2-9　基于单播前缀的 IPv6 多播地址格式

各字段的含义如下。

① 标记字段的 R 位置“0”，P、T 位则分别置“1”，表示其为基于单播前缀的多播地址。

② 保留：保留字段，必须为“0”。

③ Plen：表示网络前缀的有效长度（单位为 bit）。

④ 网络前缀：表示该多播地址所属子网的单播前缀，有效长度由 Plen 字段指定。

⑤ 组 ID：缩短为 32bit，含义不变。

（3）内嵌 RP 地址的 IPv6 多播地址

内嵌 RP 地址的 IPv6 多播地址是 IPv6 PIM 中特有的 RP 发现机制。该机制使用内嵌 RP 地址的 IPv6 多播地址，使得多播路由器可以直接从该地址中解析出 RP 的地址。

内嵌 RP 地址的 IPv6 多播地址格式如图 2-10 所示。

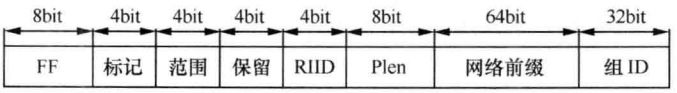


图 2-10　内嵌 RP 地址的 IPv6 多播地址格式

各字段的含义如下。

① 标记字段的 R、P 和 T 位均置“1”，表示内嵌 RP 地址的多播地址。

② 保留：保留字段，必须为“0”。

③ RIID：表示 RP 地址的接口 ID，RIID 字段填充到 RP 地址的最低 4 位。

④ Plen：表示 RP 地址前缀的有效长度（单位为 bit）。

⑤ 网络前缀：64bit。表示 RP 地址前缀，有效长度由 Plen 字段指定。

⑥ 组 ID：缩短为 32bit，含义不变。

多播 IPv6 地址对应的 MAC 地址，其高 16 位为 0x3333，低 32 位为 IPv6 多播地址的低

32 位，最终形成 48 位的多播 MAC 地址。其格式如图 2-11 所示。

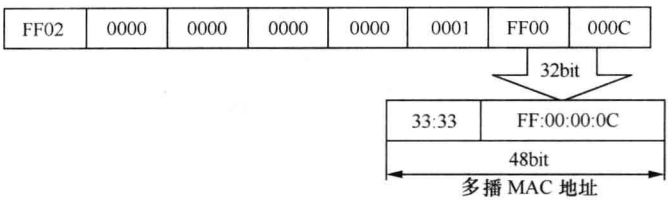


图 2-11　多播 MAC 地址格式

3．任播地址

任播地址从单播地址空间中进行分配，使用单播地址的格式，仅通过地址本身，节点无法区分任播地址和单播地址，因此配置时需要指定该地址是任播地址，且只能分配给路由器。与多播地址一样，多个节点都监听这个地址，因此任播地址只能作为目的地址使用。与多播地址不同，任播地址只有一个节点（距离本路由器最近的一个网络接口）接收 IPv6 数据分组。

在 RFC 4291 中定义了一种子网路由器任播地址（Subnet-Router Anycast Address），其标识某一子网内的所有路由器。目的地址为子网路由器任播地址的数据分组被该地址子网前缀标示的子网内的任一路由器接收，其格式如图 2-12 所示。

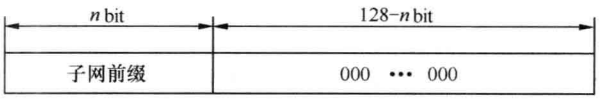


图 2-12　子网路由器任播地址格式

子网前缀为其对应的单播地址的网络前缀，用以区分或代表某一子网链路。接口 ID 值置为 0。

4．节点必须配置的地址

前面详细介绍了 IPv6 的单播、多播、任播 3 种地址类型，如果一个节点为主机，那么其必须配置的地址有哪些呢？下面一一列出。

① 每个接口都有一个链路本地地址。
② 手工或自动配置的任何单播地址。
③ 环回地址。
④ 所有节点多播地址。
⑤ 对应于单播地址的节点请求的多播地址。
⑥ 节点所属所有多播组相关的多播地址。

对于路由器而言，除上述地址以外，还有下面 4 种地址。

① 配置的用于标识一个路由器的子网路由器任播地址。
② 路由器配置的所有其他任播地址。
③ 所有路由器多播地址。
④ 该路由器所属的其他群组的多播地址。

2.2　IPv6 报文结构

IPv6 报文通常具有如图 2-13 所示的结构。

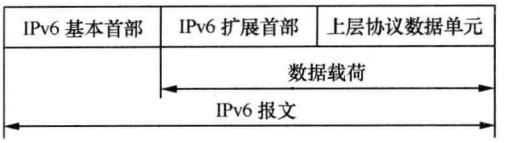


图 2-13　IPv6 报文结构图

在一个 IPv6 报文中，IPv6 基本首部是必需的，固定长度为 40Bytes。另外，基本首部之后还可以携带 0 到多个扩展首部。IPv6 协议属于网络层协议，上层协议数据单元通常是传输层的协议单元（如 UDP 和 TCP），而 ICMPv6 报文也需要 IPv6 报文来承载，但它不属于传输层。

2.2.1　IPv6 基本报头

IPv6 基本报头包含 8 个字段，总长度为 40Bytes，报头格式如图 2-14 所示。

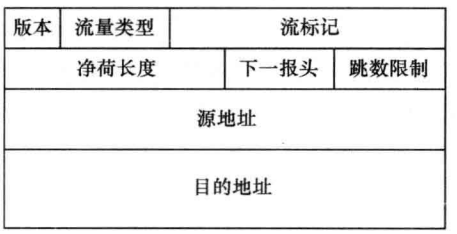


图 2-14　IPv6 基本报头

IPv6 基本报头各域的含义如下。

① 版本：其值为 6。

② 流量类型（Traffic Class）：表示 IPv6 数据分组的类型或优先级，通过此字段，信源节点和转发路由器为 IPv6 分组提供不同形式的差别服务。

③ 流标记（Flow Label）：用来标识这个数据分组属于源节点和目标节点之间的一个特定数据分组序列，它需要中间 IPv6 路由器进行特殊处理。

④ 有效负载长度（Payload length）：表示 IPv6 数据分组的有效载荷长度，即扩展报头和上层协议数据单元。

⑤ 下一报头（Next Header）：定义了紧跟基本报头之后的第一个扩展报头的类型或者上层协议数据单元中的协议类型。

⑥ 跳数限制（Hop Limit）：定义了 IPv6 分组所能经过的最大跳数，每经过一个路由器，该字段值减 1，当减为 0 时，数据分组被丢弃。

⑦ 源地址（Source Address）：表示发送方的地址。

⑧ 目的地址（Destination Address）：表示接收方的地址，在扩展报头中出现的路由报头基本报头中的目的地址并不是最终接收站的地址。

IPv6 报头比 IPv4 报头更为简洁、高效，满足新业务的发展要求。

2.2.2　扩展分组头

IPv6 中，IPv4 报头的选项内容移到了扩展分组头中，这种设计方法简化了 IPv6 报头，提高了中间路由器的转发效率。

分组投送路径上的所有节点都不检查或处理扩展报头，直到此分组到达基本报头目的地址字段指明的目的节点（并不一定是最终接收节点；在多播情况下是多播组的所有节点）后，才由目的节点处理，节点必须严格按照扩展报头在分组中出现的次序处理它们。但有一个例外，逐跳选项报头携带的信息必须被分组投送路径上的每一个节点处理。

IPv6 分组可以携带 0 个、一个或多个扩展报头，每个扩展报头都有其前面报头的下一跳值（Next Header Value）来确定。扩展报头出现的顺序是有一定原则的，其顺序如下：逐跳选项报头、目标选项报头（当存在路由报头时，用于中间目标）、路由报头、分段报头、认证报头、封装安全有效载荷报头，以及目标选项报头（用于最终目标节点）。

1．逐跳选项报头

逐跳选项报头主要用于为在传送路径上的每次跳转指定发送参数，传送路径上的每台中间节点都要读取并处理该字段。其报头结构如图 2-15 所示。

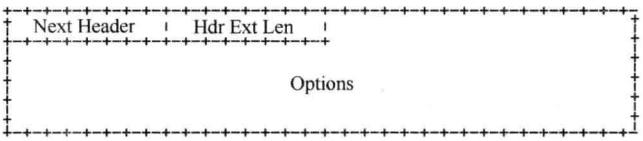


图 2-15　逐跳选项报头

各域的含义如下。

① Next Header（下一跳分组头）：长度为 8bit，表示下一个头的协议类型。

② Hdr Ext Len（报头扩展长度）：长度为 8bit，表示逐跳选项头的长度，以 8Bytes 为单位，不包括头 8Bytes。

③ Options（选项）：一系列选项字段和填充字段的组合。它可以描述数据报文转发方面的特性，也可以用作填充。一个逐跳选项报文可以包含一个或多个选项字段。

（1）选项

逐跳选项头携带了一定数量的以 TLV 方式编码的选项，结构如图 2-16 所示。

Option Type	Opt DataLen	Option Data

图 2-16　选项结构

Option Type（选项类型）：表示选项的类型，也表明了节点如何处理该选项。

Opt DataLen（选项长度）：表示选项中的字节数，其值不计入选项类型和选项字段的长度。

Option Data（选项数据）：与该选项相关的特定数据。

目前逐跳选项有以下两个重要的应用。

① 巨包的转发：在非常大的 MTU 网络上，IPv6 可以传送大于 65535Bytes 的数据分组，这就需要超大有效载荷选项（选项类型值为 194）。如果 Payload length 大于 65535Bytes，则基本报头中的有效载荷长度为 0，逐跳选项报头中的超大有效载荷选项的“选项数据”字段来表示真正的有效载荷。

② 路由器警示：IPv6 数据报中的信息希望能够得到中间路由器的查看和处理，即使这个分组是发给其他某个节点的，此时 IPv6 报文携带路由器警示选项（选项类型值为 5）。

（2）填充字段

为了符合逐跳选项报头中选项的对齐要求，通常会在选项之前进行填充。当有多个选项时，也会在两个选项之间进行填充，填充物就是 Pad1 和 Pad*N*，它们也被称为两个特别的选项。Pad1 作用是插入 1 个填充字节，而 Pad*N* 是插入 2 个或多个填充字节。

Pad1 选项结构如图 2-17 所示。

Pad1 只有 1Byte，选项类型值为 0，没有长度和填充字节。

Pad*N* 选项结构如图 2-18 所示。

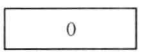


图 2-17　Pad1 选项结构

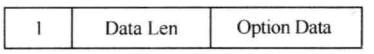


图 2-18　Pad*N* 选项结构

Pad*N* 选项包括类型字段（类型值为 1）、长度字段 Data Len（值为当前所有的填充字节数）和 0 或多个填充字节。

2．目标选项报头

目标选项报头存放用于数据分组目的地才需要处理的选项，其结构与逐跳选项报头相同，而且目标选项报头和逐跳选项报头的选项类型使用相同的编号方式。并非所有的选项都可以使用这两种报头，某些特定的选项只对应于某种选项报头。

目标选项报头有如下两种用法。

① 如果存在路由扩展头，则每一个中转路由器都要处理这些选项。

② 如果没有路由扩展头，则只有最终目的节点需要处理这些选项。

目标选项报头主要用于 MIPv6，移动节点和家乡代理之间通过该报头交换注册信息。

3．路由扩展报头

路由扩展报头用于指定报文转发必须经过的中间节点，其报头结构如图 2-19 所示。

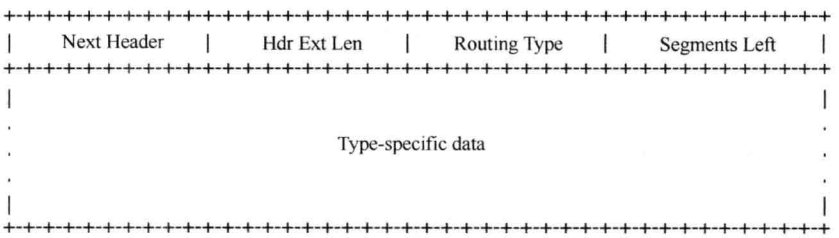


图 2-19　路由报头

各域的含义如下。

① Next Header（下一跳分组头）：表示下一个头的协议类型。

② Hdr Ext Len（报头长度）：表示路由报头的长度，不包括头 8Bytes，单位为 8Bytes。

③ Routing Type（路由类型）：定义了扩展头部的精确功能，目前定义了两种类型——类型 0 和类型 2。

④ Segments Left（剩余段数）：到达最终节点还需要经过的中间目标数。

⑤ Type-specific data（特定路由类型的数据部分）：类型 0 时，由一个 32bit 的保留字段、一个中间目标地址和最终目标地址组成的列表构成；类型 1 时，由一个 32bit 的保留字段和

一个家乡地址构成。

数据分组发送时，目标地址（基本报头）为第一个中间目标地址，路由报头的 Type-specific data 为其他中间目标地址和最终目标地址的列表（类型 1 为家乡地址），Segments Left 字段的值为 Type-specific data 中的地址总数。当数据分组被中间路由器处理时，对路由报头的处理如下。

① 基本报头的当前目标地址和路由报头地址列表中的第（*N*–Segments Left＋1）个地址相交换，*N* 为地址列表中的地址总数。

② Segments Left 字段的值减 1。

③ 转发数据分组。

需要注意的是，路由报头 Hdr Ext Len 字段的值一定是偶数，其值为地址列表个数的 2 倍。

4．分段报头

分段报头专门用于实现报文分片功能：当数据分组源节点 IP 模块发现上层协议交给它传输的数据分组大小超过了路径最大传输单元（PMTU）时，就将数据分组进行分片。

分段报头的格式如图 2-20 所示。

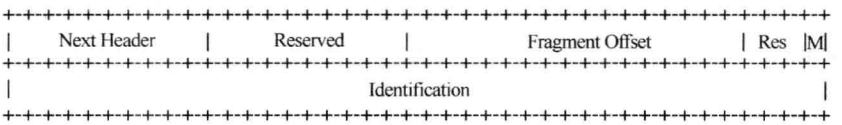


图 2-20　分段报头

各域的含义如下。

① Next Header（下一跳分组头）：表示下一个头的协议类型。

② Fragment Offset（分段偏移量）：13bit，单位为 8Bytes，表示此分组中数据的第一个字节与原来整个分组中可分段部分的数据第一个字节之间的位置关系。

③ M 标示：1bit，M 为“0”时，表示这是最后一个分段；M 为“1”时，表示后面还有分段。

④ Identification（报文标示符）： 32bit，用来唯一标识最近从源节点发送到目的节点的分组。同一个 IP 分组的不同分片具备相同的 ID。

需要注意的是，原始分组分为不可分段部分和可分段部分两部分。不可分段部分包含了 IPv6 的基本报头和必须被传送路径上的节点处理的扩展报头，即从基本报头直到路由报头（如果有的话）的所有报头，或者从基本报头直到逐跳选项报头（如果有的话）的所有报头，或者仅为基本报头；可分段部分包含仅能被最终目的节点处理的扩展报头以及上层报头和数据。

5．认证报头

认证报头能对 IP 报头和高层协议数据提供认证，实现了数据完整性和对数据来源的认证，并提供反重放保护（可选，只有当接收点验证序列号时该业务才有效）。完整性保证了数据分组在传输过程中没有被篡改过；数据分组来源的认证，保证了数据确实来自于源地址所标识的接口。

认证报头格式如图 2-21 所示。

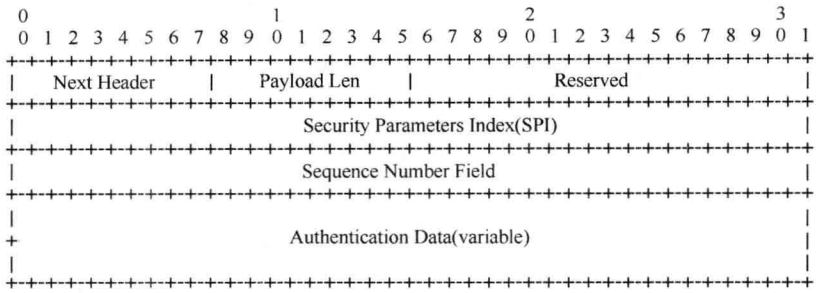


图 2-21　认证报头

各域的含义如下。

① Next Header（下一跳分组头）：表示下一个头的协议类型。

② Payload Len（报头长度）：8bit，单位为 8Bytes，表明认证报头的长度。

③ Security Parameters Index（SPI，安全系数索引）：32bit，其值可以是任意的。SPI 值和目的 IP 地址以及安全协议，唯一地表明数据分组的安全关联，0～255 为保留值。

④ Sequence Number Field：序列号字段，发送者必须发送，接收者可以完全按照此字段执行，也可以不遵照此字段执行，如果接收端开启反重播放服务，则接收端验证序列号。

⑤ Authentication Data：认证数据，可变长字段，包含有分组的整体检验值，这个字段的值必须是 32 的整数倍，因此可能包含显式的填充。

认证报头有两种应用方式：传输模式和隧道模式。

（1）传输模式

传输模式仅适用于主机程序。在 IPv6 中，认证报头被当作端到端的载荷，出现在逐跳选项报头、路由报头和分段报头之后，目标选项报头之前或之后（取决于语义的要求）。被认证部分为目标选项报头、高层协议数据和其余 IP 报头不可变部分。认证报头的位置如图 2-22 所示。

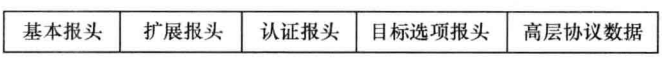


图 2-22　传输模式

（2）隧道模式

隧道模式下认证报头可以应用于主机或者安全网关上。在隧道模式下，下层的 IP 报头携带最终的源地址和目的地址，而上层的 IP 报头可以包含其他的地址，如安全网关的地址。认证报头对包括 IP 报头在内的整个下层的 IP 分组提供保护。在隧道模式下，认证报头相对于上层的 IP 报头的位置与传输模式下一样。认证报头的位置如图 2-23 所示。

图 2-23　隧道模式

隧道模式下，除了可变部分，认证报头对分组所有部分都进行认证。

6．封装安全有效载荷

封装安全有效载荷（ESP）报头和报尾提供数据机密性、数据验证、数据完整性以及防

重放服务。ESP 只对高层数据提供保护，而不对 IP 报头字段提供任何保护，除非这些字段被封装安全载荷封装（IP 报头成为某个数据分组的载荷）。

封装安全有效载荷报头在 IPv6 和 IPv4 中提供混合的安全业务，封装安全有效载荷机制可以与认证报头一起应用，或者单独在隧道模式下以嵌套的方式应用。

ESP 格式如图 2-24 所示。

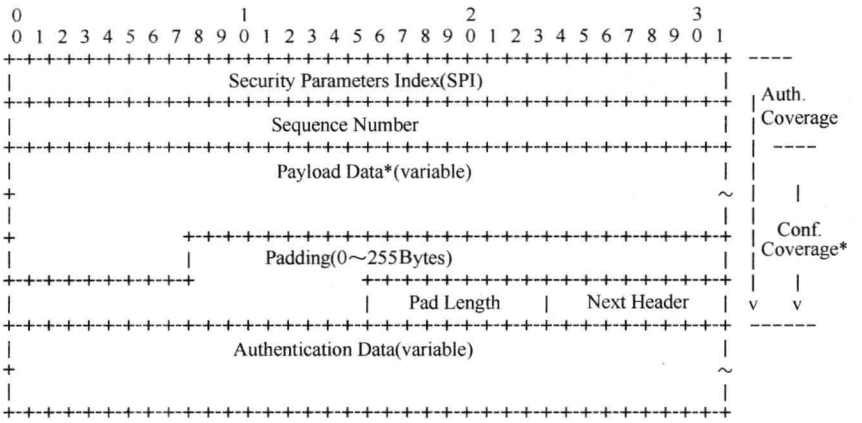


图 2-24　ESP 格式

各域的解释如下。

① Payload Data（载荷数据）：其长度是字节的整数倍，包含由 Next Header 字段描述的数据。

② Padding（扩展）：目的是为了加密、对齐和隐藏载荷的真实长度。

③ Authentication Data（认证数据）：可变长字段，包含有分组的整体检验值。这个字段是可选的，只有当安全关联包含了认证业务时才出现。

ESP 报头有两种应用方式：传输模式和隧道模式。

（1）传输模式

传输模式仅适用于主机应用程序。ESP 报头被当作端到端的载荷。它出现在逐跳选项报头、路由报头和分段报头之后，目标选项报头之前或之后（取决于语义的要求）。认证报头的位置如图 2-25 所示。

基本报头	扩展报头	封装安全载荷	目标选项报头	高层协议数据	封装安全载荷报尾	封装安全载荷认证

图 2-25　ESP 传输模式

图 2-25 所示的分组中，被加密的部分是目标选项报头、高层协议数据和封装安全载荷报尾。而被认证的部分除了有被加密的部分之外，还要加上封装安全载荷。

（2）隧道模式

隧道模式下认证报头可以应用于主机或者安全网关上。在隧道模式下，下层的 IP 报头携带最终的源地址和目的地址，而上层的 IP 报头可以包含其他的地址，如安全网关的地址。隧道模式下，ESP 报头对包括 IP 报头在内的整个下层的 IP 分组提供保护。在隧道

模式下，ESP 报头相对于上层的 IP 报头的位置与传输模式下一样。ESP 报头的位置如图 2-26 所示。

上层基本报头	上层扩展报头	封装安全载荷	下层基本报头	下层扩展报头	高层协议数据	封装安全载荷报尾	封装安全载荷认证

图 2-26　ESP 隧道模式

图 2-26 所示的分组中，被加密的部分是下层基本报头、下层扩展报头、高层协议数据和封装安全载荷报尾。而被认证的部分除了有被加密的部分之外，还要加上封装安全载荷。

2.3　ICMPv6 协议

ICMPv6 是一种差错和控制报文协议，位于 TCP/IP 模型中的网络层，它不仅用于传输差错报文，还用于传输控制报文。ICMPv6 主要用于报告在转发 IPv6 数据分组过程中出现的错误，并提供诊断功能和附加的主机功能。ICMPv6 报文使用 IPv6 报文承载，其消息类型又分为两类——错误消息和信息消息。

2.3.1　ICMPv6 报文结构

ICMPv6 报文结构如图 2-27 所示。

各域的含义如下。

① 类型：8bit，表明 ICMPv6 报文的类型。最高位为 0 时，表示该报文为差错报文；最高位为 1 时，表示该报文为信息报文。

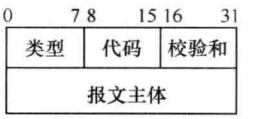


图 2-27　ICMPv6 报文结构

② 代码：8bit，依赖于“类型”域的值。在特定类型之下，定义了下一层的 ICMPv6 报文类型，指明其具体功能。

③ 校验和：16bit，检验 ICMPv6 报文和部分 IPv6 报头数据的正确性。

④ 报文主体：其长度和内容依据报文类型不同而不同。

2.3.2　ICMPv6 报文类型

ICMPv6 报文分为差错报文和信息报文。

1．差错报文

差错报文主要包括目的地不可达报文、分组太长报文、超时报文和参数错误报文。

（1）目的地不可达报文

这个报文由路由器或源主机在由于除业务流拥塞之外的原因而无法转发一个分组的时候产生。

（2）分组太长报文

当接收某分组的路由器由于分组长度大于将要转发到的链路的 MTU，而无法对其进行转发时，将会产生分组太长报文。该 ICMPv6 错误报文中有一个字段指出导致该问题的链路的 MTU 值。在路径 MTU 发现过程中这是一个有用的错误报文。

（3）超时报文

当路由器收到一个跳数极限为 1 的分组时，它必须在转发该分组之前减小这个数值。如果在路由器减小该数值后，跳数极限字段的值变为 0（或者是路由器收到一个跳数限制字段为 0 的分组），那么路由器必须丢弃该分组，并向源节点发送 ICMPv6 超时报文。源节点在收到该报文后，可以认为最初的跳数限制设置得太小（分组的真实路由比源节点想象的要长），也可以认为有一个选路循环导致分组无法交付。

（4）参数错误报文

当 IPv6 报头或扩展头中的某些部分有问题时，路由器由于无法处理该分组而会将其丢弃。路由器的实现中应该可以产生一个 ICMP 参数错误报文来指出问题的类型（如错误的报头字段、无法识别的下一个报头类型或无法识别的 IPv6 选项），并通过一个指针值指出在第几个字节遇到这种错误情况。

2．信息报文

ICMPv6 信息报文包括回声请求报文和回声应答报文。回声请求报文可以向任何一个正确的 IPv6 地址发送，并在其中包含一个回声请求标识符、一个顺序号和一些数据。尽管二者都是可选项，但回声请求标识符和顺序号可以用来区分对应不同请求的响应，回声请求的数据也是一个选项，并可用于诊断。

当一个 IPv6 节点收到一个回声请求报文后，它必须回送一个回声应答报文，并在应答中包含相同的请求标识符、顺序号和在最初的请求报文中携带的数据。ICMP 回声请求/应答报文对是 Ping 功能的基础。Ping 是一个重要的诊断功能，因为它提供了一种方法来决定一个特定的主机是否与其他一些主机连接在相同的网络上。

3．ICMPv6 报文信源地址

发送 ICMPv6 报文的节点必须在计算校验和之前确定 IPv6 报头中的 IPv6 信源和信宿地址。如果节点有多个单播地址，它必须按照以下规则选择报文的信源地址。

① 如果报文是对送到节点的某个单播地址的报文的回复响应，则此响应的信源地址必须是该节点的同一个单播地址。

② 如果报文是对送到节点所在的多播组或任播组的报文的回复响应，则此响应的信源地址必须是一个单播地址，该单播地址属于多播或任播数据分组被接受的接口。

③ 如果报文是对送到不属于节点地址的报文的回复响应，则此响应的信源地址必须是该节点的单播地址。该单播地址应该对于诊断差错最有帮助。节点的路由表必须经过检查，以决定哪个接口被应用于将报文发送到信宿，属于该接口的单播地址必须作为信源地址。

需要注意的是，ICMPv6 报文封装在 IPv6 数据分组的数据部分进行传输，但并不把 ICMPv6 看作比 IPv6 更高的协议，ICMPv6 报文的信宿软件作为 IP 软件的一模块而存在。

2.4　邻居发现技术

邻居发现协议（Neighbor Discovery Protocol，NDP）使用 ICMPv6 报文来实现相邻节点（同一链路上的节点）的交互管理，并在一个子网中保持网络层地址和链路层地址之间的映射。

它取代了 IPv4 协议中使用的地址解析协议（ARP）、Internet 控制报文协议（ICMP）、路由器发现以及 ICMP 重定向报文，并提供额外的功能。

2.4.1 ND 概述

邻居发现协议中定义了 5 种类型的报文：邻居请求、邻居通告、路由器请求、路由器通告和重定向。

1. 路由器请求报文

主机发送一个路由器请求报文来提示路由器尽快创建一个路由器通告报文立即发出通告，而不必等待下一个预定传送时间到达。主机接收到路由器通告报文后，将其丢弃。

路由器请求报文格式如图 2-28 所示。

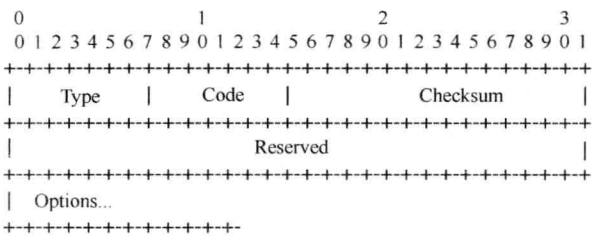


图 2-28 路由器请求报文格式

对路由器请求报文各域的说明如下。

（1）IP 域

① 源地址：发送接口的地址，或者未指定地址。

② 目的地址：全路由器多播地址（FF02::2）。

③ 跳数限制：值为 255。

④ 认证头：如果在发送者和目的地址之间存在对认证头的安全关联，那么发送者必须包含该头。

（2）ICMP 域

① Type（类型）：8bit，值为 133。

② Code（代码）：8bit，值为 0。

③ Checksum（校验和）：16bit，ICMP 校验和。

④ Reserved（保留字）：32bit，发送者必须将其初始化为 0，接收者必须将其忽略。

⑤ Options（选项）：源链路层地址。如果源地址是未指定地址，那么一定不能包含该选项，否则，则应该包含该选项。

2. 路由器通告报文

路由器周期性产生路由器通告报文来告知网络节点本路由器的可用性，以及用于配置的链路和 Internet 参数。路由器也使用路由器通告报文对节点发出的路由器请求报文做出响应。路由器通告报文包含对所使用的网络地址前缀、建议的逐跳极限值及本地 MTU 的指示，也包括指明节点应使用的自动配置类型的标志。路由器通过多播来发送其路由器通告报文。这样，同一链路上的节点可以构造自己的可用默认路由器列表。路由器通告报文格式如图 2-29 所示。

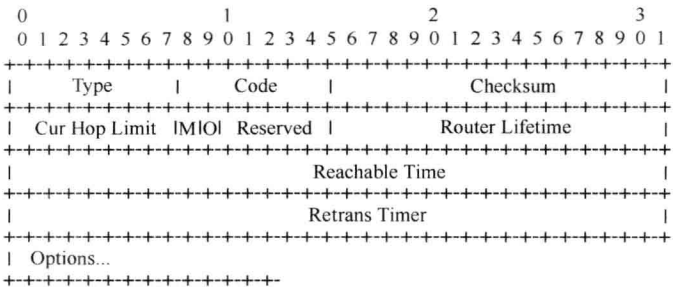


图 2-29　路由器通告报文格式

对路由器通告报文各域的说明如下。

（1）IP 域

① 源地址：发送接口的链路本地地址。

② 目的地址：路由器请求的源地址，或者全节点多播地址（FF02::1）。

③ 跳数限制：值为 255。

④ 认证头：如果在发送者和目的地址之间存在对认证头的安全关联，那么发送者必须包含该头。

（2）ICMP 域

① Type（类型）：8bit，值为 134。

② Code（代码）：8bit，值为 0。

③ Checksum（校验和）：16bit，ICMP 校验和。

④ Cur Hop Limit（当前跳数限制）：8 位无符号整数，对于要发出的 IP 分组的 IP 头的 Hop Count 域，默认值应该设置为“当前跳数限制”，如果该值为 0，则标识该值未被当前路由器指定。

⑤ M 标志：标识是否使用有状态地址配置技术。

⑥ O 标志：标识是否有非地址需要有状态协议配置。

⑦ Reserved（保留字）：6bit。发送者必须将其初始化为 0，接收者必须将其忽略。

⑧ Router Lifetime（路由器生存期）：16 位无符号整数。为 0 表示该节点不是默认路由器，不应该在默认路由器列表中出现。

⑨ Reachable Time（可达时间）：32 位无符号整数，用于主机判断邻居是否可达。

⑩ Retrans Timer（重传定时器）：32 位无符号整数，用于触发重发邻居请求报文。

⑪ Options：选项。

a．源链路层地址：发送路由器通告报文接口的链路层地址。

b．最大传输单元：在拥有可变最大传输单元的链路上，该选项必须传送。

c．前缀信息：指明联网和用于自动地址配置的前缀。

3．邻居请求报文

节点发送邻居请求报文来请求邻居的链路层地址，以验证它先前所获得并保存在高速缓存中的邻居链路层地址的可达性，或者验证它自己的地址在本地链路上是否唯一。

邻居请求报文格式如图 2-30 所示。

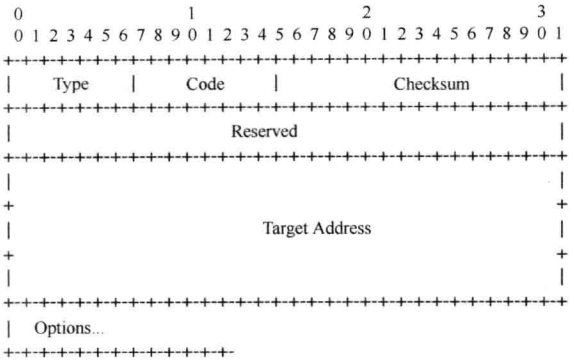


图 2-30　邻居请求报文格式

对邻居请求报文各域的说明如下。

（1）IP 域

① 源地址：发送接口的地址，或者未指定地址。

② 目的地址：与目标地址相对应的被请求节点多播地址（用于解析 IPv6 地址时），或就是目标单播地址（用于邻居的可达性监测时）。

③ 跳数限制：值为 255。

④ 认证头：如果在发送者和目的地址之间存在对认证头的安全关联，那么发送者必须包含该头。

（2）ICMP 域

① Type（类型）：8bit，值为 135。

② Code（代码）：8bit，值为 0。

③ Checksum（校验和）：16bit，ICMP 校验和。

④ Reserved（保留字）：32bit，发送者必须将其初始化为 0，接收者必须将其忽略。

⑤ Target Address：（目标地址）：128bit，请求目标的地址，不能是多播地址。

⑥ Options（选项）：源链路层地址。如果源地址是未指定地址，那么一定不能包含该选项；否则，则应该包含该选项。

4．邻居通告报文

节点在收到邻居请求报文的请求或其链路层地址改变时，发出邻居通告报文。

邻居通告报文格式如图 2-31 所示。

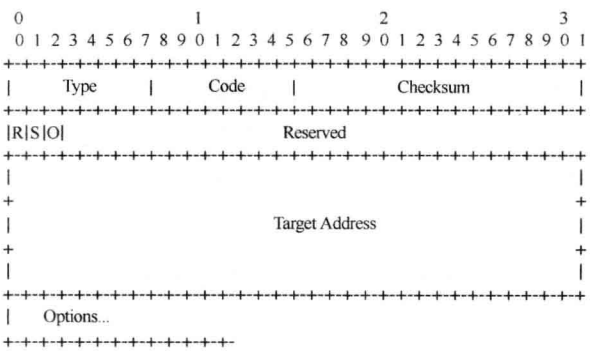


图 2-31　邻居通告报文格式

对邻居通告报文各域的说明如下。

（1）IP 域

① 源地址：发送接口地址。

② 目的地址：邻居请求的源地址，或者该地址所在的被请求节点多播组，或者全节点多播地址（FF02::1）。

③ 跳数限制：值为 255。

④ 认证头：如果在发送者和目的地址之间存在对认证头的安全关联，那么发送者必须包含该头。

（2）ICMP 域

① Type（类型）：8bit，值为 136。

② Code（代码）：8bit，值为 0。

③ Checksum（校验和）：16bit，ICMP 校验和。

④ R 标志：该位为“1”，表示发送者为路由器。

⑤ S 标志：该位为“1”，表示该通告是针对响应邻居请求报文的。

⑥ O 标志：标识是否覆盖主机相关表项信息。

⑦ Reserved（保留字）：29bit。发送者必须将其初始化为 0，接收者必须将其忽略。

⑧ Target Address（目标地址）：128bit，不能是多播地址。

⑨ Options（选项）：目标链路层地址，目标地址对应接口的链路层地址。

5. 重定向报文

路由器发送重定向报文以通知主机，对于特定目的地更好的下一跳地址。主机可以据此定位到一个更好的下一跳路由器，也可以通过该报文获悉，目的地址其实就是邻居节点。

重定向报文格式如图 2-32 所示。

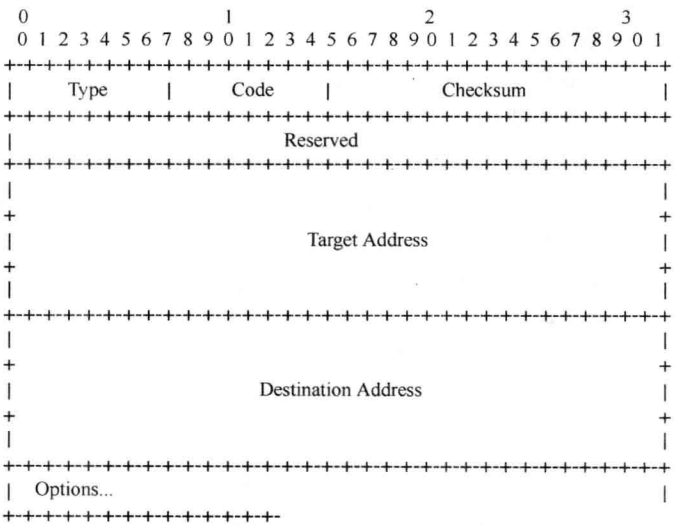


图 2-32　重定向报文格式

对重定向报文各域的说明如下。

（1）IP 域

① 源地址：发送接口的链路本地地址。

② 目的地址：触发重定向事件的分组的源地址。

③ 跳数限制：值为 255。

④ 认证头：如果在发送者和目的地址之间存在对认证头的安全关联，那么发送者必须包含该头。

（2）ICMP 域

① Type（类型）：8bit，值为 137。

② Code（代码）：8bit，值为 0。

③ Checksum（校验和）：16bit，ICMP 校验和。

④ Reserved（保留字）：32bit。发送者必须将其初始化为 0，接收者必须将其忽略。

⑤ Target Address（目标地址）：128bit，不能是多播地址，用于 ICMP 目的地址更好的下一跳地址。当目标是一个实际端点时，目标地址和目的地址的值相同；否则，应该是更好的下一跳路由器的链路本地地址。

⑥ Destination Address（目的地址）：要被重定向到目的 IP 地址。

⑦ Options：选项。

a．目标链路层地址：目标地址对应接口的链路层地址。

b．被重定向头：尽可能多地发出重定向事件的 IP 数据分组，但重定向报文不能超过 1280 个 8 位组。

2.4.2 ND 功能

ND 协议可以实现诸多功能，包括无状态自动地址配置、重复地址检测、地址解析、邻居不可达检测以及重定向等。

1．无状态自动地址配置

无状态自动地址配置技术基于 IPv6 地址结构，一个 IPv6 地址由前缀和接口 ID 组成。实现动态配置就是实现这两个部分的动态配置。

接口 ID 的长度为 64bit。获取 IPv6 接口 ID 最常用的方法是使用 IEEE EUI-64 地址。EUI-64 通过 MAC 地址生成接口 ID，这个部分在后面的章节有详细介绍。

IPv6 地址前缀的实际作用是标识主机与路由器之间的网络，所以主机需要的前缀就是路由器接口的前缀，为了自动获得这个前缀，在路由器和主机之间运行一个无需配置主机的协议即可。这里需要用到 ND 协议的路由器请求和路由器通告消息，前者用于发现路由器，并促使路由器发送路由器通告消息通报前缀信息。

前缀信息是作为 ICMP 选项的形式承载在报文中的，前缀选项的格式如图 2-33 所示。

对各主要字段解释如下。

① Prefix Length（前缀长度）：指示 Prefix 的长度。

② L：表示该前缀是否可以用于判断一个地址是否在本地链路上。

③ A：表示该前缀是否可以用于地址自动配置。

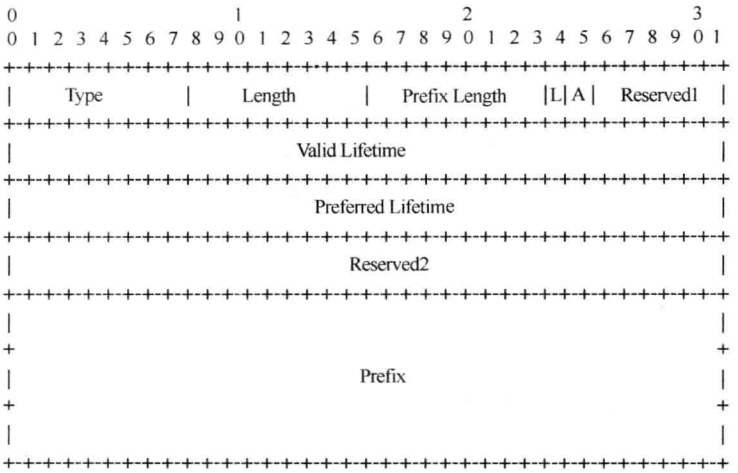


图 2-33　前缀信息选项格式

④ Valid Lifetime（有效生命期）和 Preferred Lifetime（优先生命期）：用于表示前缀的两个生命期。对于一个在 Preferred Lifetime 内的前缀生成的地址，任何上层应用都可以不受限制地使用它；对一个超过 Preferred Lifetime 但未超过 Valid Lifetime 内的前缀生成的地址，正在使用此地址的上层应用还可以继续使用，但新的上层应用不能使用这个地址；对于一个超过 Valid Lifetime 的前缀生成的地址，任何上层应用都不能使用它。

需要注意的是，这个路由器通告报文是发给所有节点多播组的。这是因为路由器请求报文不仅仅只配置一台主机。主机收到这个信息以后，就可以配置自己的 IP 地址了。

2．重复地址检测

IPv6 重复地址检测应用 ND 协议的邻居请求消息和邻居通告消息。ND 协议的传输协议——ICMPv6 工作在第三层。可以利用三层多播技术，限制重复检测报文的传输范围，仅需发送到待检测地址所在多播组即可。

一个地址分配给一个接口之后通过重复地址检测之前被称为 tentative 地址，该接口加入所有节点多播组和该地址对应的被请求节点多播组。

节点向一个 tentative 地址所在的多播组发送一个邻居请求报文（Target 域包含这个 tentative 地址）。如果收到某个其他站点回应的邻居通告报文，则证明该地址在网络上是重复的；如果在尝试一定次数后，没有收到相关的邻居通告报文，那么就启用该地址。

3．地址解析

在仅仅知道邻居节点 IP 地址的情况下，节点通过邻居请求报文和邻居通告报文，获得邻居的链路层地址的过程，就是地址解析。

节点的地址解析由多播发送邻居请求报文开始。该请求报文要求目标节点返回它的链路层地址。源路由器在邻居请求报文中包含了它自己的链路层地址，并将邻居请求报文以多播的方式发送到与目标地址相关的多播地址上；目标路由器在单播邻居公告报文中返回它自己的链路层地址。

4．邻居不可达性检测

（1）邻居状态

IPv6 通信技术基本就是建立、维护与使用路由表、邻居缓存这两个数据结构的技术。邻

居缓存条目有 5 种状态，即：不完全、可达、陈旧、延迟和探测。

① 不完全（Incomplete）：正在进行地址解析，邻居的链路层地址还没确定。

② 可达（Reachable）：邻居在最近处于可达状态（在小于 10s 以内）。

③ 陈旧（Stale）：在数据流发送给邻居以前该邻居是不可达的，并无法验证其可达性。

④ 延迟（Delay）：邻居不再是可达的，同时数据流在最近已经发送给邻居，但不立即对该邻居进行探测，而在一个短时延后发送探测信息，这样就可以为上层协议提供可达性确认。

⑤ 探测（Probe）：邻居不再是可达的，同时发送单播邻居请求探测以验证其可达性。

其状态如图 2-34 所示。

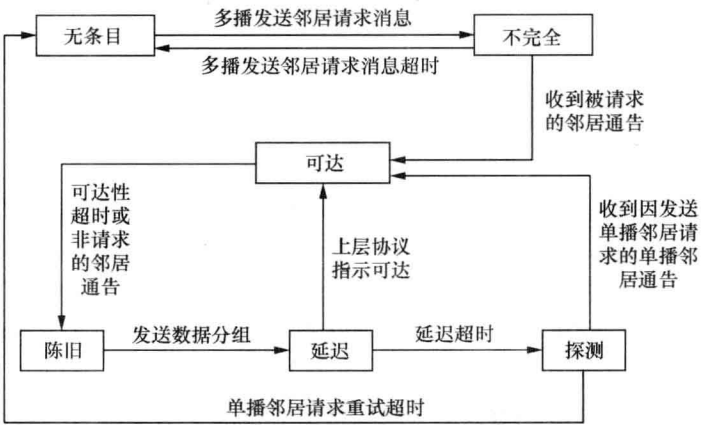


图 2-34　邻居状态

（2）邻居不可达性检测

节点通过邻居不可达性检测来确定以前与其相连的节点现在是否依然连通。

为了更具体地观察邻居不可达检测所起的作用，下面给出 PC1 和 PC2 长期互相不通信之后，PC2 再次发起通信时，PC1 上关于 PC1 邻居表项的实际变化过程。

在搁置一段时间之后，PC1 邻居表项中关于 PC2 的条目状态变为“陈旧”。现在主动让 PC2 向 PC1 发 Ping 分组，PC1 根据处于“陈旧”状态的邻居条目（关于 PC2 的）封装报文，回送 PC2，此时该邻居条目进入“延迟”状态。由于在“延迟”状态并未获得 PC2 的可达性信息，PC1 邻居表项中关于 PC2 的条目最终进入探测状态，PC1 发送邻居请求报文来探测 PC2 的可达性，PC2 收到这个报文后回应邻居通告报文，并导致 PC1 改变状态为“可达”。

5. 重定向

当目的节点实际就是邻居节点时，发送给它的数据报便不需要路由器转发，路由器通过重定向报文即可告知主机；当分组必须发送到一个非连接的目的地时，需要选择转发分组的路由器；当选择的路由器作为消息传送的下一跳而并不是最好的下一跳时，路由器需产生重定向消息，通知源节点到达目的地存在一个更佳的下一跳路由器。路由器必须能够确定每个邻居路由器的本地链路地址，以保证重定向消息中的目标地址根据本地链路地址来识别邻居路由器。在源端没有正确应答重定向消息，或者源端选择忽略没有被验证的重定向消息的情况下，为了节省带宽和处理的开销，路由器必须限定发送重定向消息的速率。

2.5　IPv6 路由协议

随着 IPv6 时代的来临，目前广泛应用的 IPv4 路由协议需要进行扩展或重新设计以提供对 IPv6 技术的支持。下面，我们对 IPv6 单播路由协议（IGP、EGP）和 IPv6 多播协议进行介绍。

2.5.1　IPv6 单播路由协议

IPv6 单播路由协议包括内部网关协议（IGP）和外部网关协议（External Gateway Protocol，EGP）两种。IPv6 IGP 主要的有 RIPng、OSPFv3、IS-ISv6，IPv6 EGP 一般是指 BGP4 + 。

1. RIPng

RIPng 是在 RIP 基础上开发的用于 IPv6 网络的路由协议，在工作机制上与 RIP 基本相同，是 IPv6 中基于距离向量的内部网关路由协议。但为了支持 IPv6 地址格式，RIPng 对 RIP 做了一些改动。RIPng 一般作为中等或者偏小规模的网络自治系统中的内部网关路由协议。

（1）RIPng 技术实现

RIPng 通过 UDP 报文进行路由信息的交换，使用端口号 521 发送和接收数据报。特别的查询信息可以不从源节点端口 521 发出，但是必须发送到目标节点的端口 521。

每个路由器都有接口连接一个或者多个网络（直连网络）。RIPng 协议的实现依赖这些网络的相关信息，包括目的地址前缀、前缀长度以及度量等。RIPng 使用跳数（hop count）作为度量（metric）。RIPng 网络的度量是 1～15 之间的整数，大于或等于 16 的跳数定义为无穷大，即目的网络或主机不可达。

RIP 的启动和运行过程如下。

① 路由器启动 RIP 后，需要尽快建立起路由表。路由器请求邻居路由器全部路由表，请求报文以多播方式从 RIPng 端口（端口 521）发出。

② 路由器收到 RIPng 请求报文后，将自身全部路由表（注意水平分割）单播发往请求报文的源地址和端口。

③ 路由器收到响应报文后，将处理响应报文中的路由表项（RTE）并建立和更新自己的路由表，同时向相邻路由器多播发送触发更新报文，通告路由更新信息。

④ 相邻路由器收到触发更新报文后，又向其各自的相邻路由器发送触发更新报文。在一连串的触发更新后，各路由器都能得到并保持最新的路由信息。

⑤ RIPng 周期性地向相邻路由器多播发送本地路由表，同时采用老化机制对超时的路由进行老化处理，保证路由的实时性和有效性。

RIPng 工作机制与 IPv4 RIP 基本相同，实现简单，主要适用于中小规模网络。

（2）RIPng 与 RIP 的不同点

为了适应 IPv6 网络环境，RIPng 对 RIP 进行了改进，主要体现在以下几方面。

① 报文格式不同。RIPng 有两类 RTE：目的前缀 RTE 和下一跳 RTE。目的前缀 RTE 指明可达目的网络，下一跳 RTE 为 RIPng 提供了直接指定下一跳 IPv6 地址的能力。下一跳 RTE 指明的 IPv6 地址适用于跟随其后的目的前缀 RTE，直到 RIPng 报文结束或者出现另一个下一跳 RTE 为止。

② 报文长度不同。RIPv2 规定每个报文最多只能携带 25 个 RTE；RIPng 对报文长度、RTE 的数量不作规定，报文的长度由邻居路由器之间链路 MTU 决定。

③ 发送方式不同。RIP 可以根据用户配置采用广播或多播方式来周期性地发送路由信息；RIPng 使用多播方式周期性地发送路由信息。

④ 端口号不同。RIP 使用的端口号 520 发送和接收数据报；RIPng 使用端口号 521 发送和接收数据报。

⑤ 安全机制不同。RIP 使用认证 RTE 作为安全机制；RIPng 自身不提供认证功能，依赖 IPv6 提供的安全机制，保证 RIPng 报文的完整性和机密性。

2．OSPFv3

OSPFv3 是在 OSPFv2 基础上开发的用于 IPv6 网络的路由协议。OSPFv3 的机制和选路算法并没有本质的改变，其处理流程与 OSPFv2 基本保持一致。为了支持 IPv6 地址格式，OSPFv3 在 OSPFv2 的基础上进行了一些改进。

（1）OSPFv3 技术实现

为了支持在 IPv6 环境中运行，OSPFv3 对 OSPFv2 做出了一些必要的改进，但作为链路状态路由选择算法，其实现机制没有本质改变。

OSPFv3 运行在 IPv6 网络中，它同 OSPFv2 并不兼容，但 OSPFv3 的处理流程与 OSPFv2 基本保持一致，如泛洪过程，DR 选举。对区域的支持以及 SPF 的计算流程，OSPFv3 只是在 OSPFv2 的基础上进行了一些改进，以支持报文格式的变化并处理 IPv6 中 128bit 的地址。

（2）OSPFv3 与 OSPFv2 的不同点

OSPFv3 在 OSPFv2 基础上做出了一些必要的改进，这些改进主要包括以下几方面。

① 链路概念取代网络概念

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 的实现是基于链路的，同一链路不同子网上的节点也可以直接通话。

② 报文去除地址语义

对于 OSPFv3 来说，除了 LS Update 报文载荷中存在地址以外，协议报文中不再提供地址信息；Router-LSA 和 Network-LSA 中也不再包含网络地址；OSPF Router ID，区域 ID 和 Link State ID 仍然保留 IPv4 中 32bit 的长度，因此不能使用 IPv6 地址来代表这些信息。

③ 增加泛洪的范围

LSA 的泛洪范围已经被明确地定义在 LSA 的 LS Type 字段，目前有以下 3 种 LSA 泛洪范围。

☆ 本链路范围：用于 Link LSA。

☆ 区域范围：用于 Router LSA、Network LSA、Inter Area Prefix LSA、Inter Area Router LSA 和 Intra Area Prefix LSA。

☆ 自治域范围：用于 AS-external-LSA。

④ 链路支持多实例复用

OSPFv3 通过在 OSPF 报文中增加一个 Instance ID 字段实现链路复用并节约成本。

⑤ link-local 地址的使用

link-local 地址只能用于 Link LSA，而不允许用于其他类型 LSA 中。

⑥ 认证方式变化

IPv6 中提供了专门的认证报头和封装安全有效载荷报头来提供身份验证，去除了 OSPF 报文中专门的认证字段。

⑦ OSPF 报文和 LSA 格式变化

与 OSPFv2 报文头相比，OSPFv3 报文头长度只有 16Bytes，去掉了认证字段但增加了 Instance ID 字段；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；与 OSPFv2 报文头相比，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；OSPFv3 新增了 Link LSA 和 Intra Area Prefix LSA。

⑧ 对未知类型 LSA 的处理机制

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。

⑨ 只根据 Router ID 来识别邻机节点

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；在 OSPFv3 中，对于任何网络类型，都通过 Router ID 来唯一标识邻居。

3. IS-ISv6

为了能在 IPv6 环境中运行，IS-ISv6 对 IS-IS 进行了简单的扩展，使得其能够处理 IPv6 的路由信息。IS-ISv6 可以同时承载 IPv4 和 IPv6 的路由信息，完全可以独立用于 IPv4 网络和 IPv6 网络。

为了有效地支持 IPv6，IS-ISv6 新添加了支持 IPv6 路由信息的两个 TLV（Type-Length-Values）和一个新的 NLP ID（Network Layer Protocol Identifier）。TLV 是在 LSP（Link State PDUs）中的一个可变长结构，新增的两个 TLV 分别介绍如下。

（1）IPv6 Reachability（TLV Type 236）

类型值为 236（0xEC），对应于 IS-IS 中的普通可达性 TLV 和扩展可达性 TLV，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。

（2）IPv6 Interface Address（TLV Type 232）

类型值为 232（0xE8），对应于 IPv4 中的 IP Interface Address TLV，只不过把原来 32bit 的 IPv4 地址改为 128bit 的 IPv6 地址。

NLP ID 是标识 IS-IS 支持何种网络层协议的一个 8bit 字段，IPv6 对应的 NLP ID 值为 142（0x8E）。当路由器支持 IS-ISv6 功能时，那么它必须在 Hello 报文中携带该值向邻居通告它支持 IPv6。

4. 多拓扑 IS-ISv6 实现技术

使用 IS-ISv6 实现 IPv6 扩展的前提是所有的 IPv6 和 IPv4 拓扑信息必须一致。如果网络中一些路由器或者链路不支持 IPv6，从而导致 IPv6 和 IPv4 拓扑的不同，而支持双协议栈的路由器感知不到这种情况，IPv6 数据流仍会被转发到这些不支持 IPv6 的路由器或链路，转发的数据流就会被丢弃处理。IS-IS MT 用来解决以上提到的问题，在 IPv6 独立的拓扑上运行 SPF 算法。

每个 MT 运行自己的计算过程，而在计算期间，虚拟节点的 LSP 被所有拓扑使用。每个 MT 拓扑使用本拓扑的 ATT 和 OL 标志，该标志位于本拓扑的 MT TLV 中。SPF 的反向连接检查必须跟随正确的 MT 路径，以保证双向可达性。每个计算结果必须保存在隔离的 RIB 中，否则不同拓扑的重叠地址将引起不必要的路由行为。部署和配置必须保证 IP 报文只穿越定义好的节点和链路。需要一些额外的机制去选择正确的 RIB，以保证进入的 IP 报文选择正确的 RIB 并进行正确的转发。

5. BGP4 +

BGP4 + 只是在 BGP4 的基础上进行了一些扩展，原有的消息机制和路由机制并没有改变。BGP4 + 同 BGP4 完全兼容，可以独立地在 IPv4 网络上或者是 IPv6 网络上运行。

BGP4 + 使用了一个 BGP 的特殊的属性 MP-BGP 来承载 IPv6 的路由信息，这种路由信息被称为 IPv6 NLRI。BGP4 + 中引入了两个 NLRI 属性。

① MP_REACH_NLRI：Multi-protocol Reachable NLRI，多协议可达 NLRI；用于发布可达路由及下一跳信息。

② MP_UNREACH_NLRI：Multi-protocol Unreachable NLRI，多协议不可达 NLRI；用于撤销不可达路由。

BGP4 + 中的 Next_Hop 属性用 IPv6 地址来表示，可以是 IPv6 全球单播地址，也可以是下一跳的链路本地地址。

6. 路由协议选择

外部网关协议（EGP）只有一种选择，即 BGP4 + 。内部网关协议（IGP）的选择需要根据各种 IGP 特点以及网络特点进行综合考虑。小规模网络可以选择 RIPng 作为其 IGP，而大规模网络和骨干网通常选择 IS-ISv6 或 OSPFv3 作为其 IGP。

双协议栈网络 IGP 的选择有以下几种。

（1）OSPFv2 + OSPFv3

原有网络使用 OSPFv2 时，该选择较常见。IPv4 和 IPv6 可以遵循不同的拓扑，OSPFv3 进程失败不会影响 OSPFv2，反之亦然。但这种方式存在两个进程占用路由器 CPU 资源较多的缺陷。

（2）IS-ISv4/IS-ISv6

原有的网络使用 IS-ISv4 时，该选择较常见。IS-ISv6 在路由器中为一个进程占用路由器 CPU 资源较少，但进程的崩溃将会造成 IPv4/IPv6 网络同时崩溃，而且如果采用单拓扑 IS-ISv6，要求 IPv4/IPv6 路由拓扑必须相同。

（3）OSPFv2 + IS-ISv6

原有网络使用 OSPFv2，目前使用的路由器不能很好地支持 OSPFv3 时，作为过渡期，有时会用这种选择，但 IPv6 only IS-IS 并不十分稳定，互操作性也较差。

（4）IS-ISv4 + OSPFv3

这种选择很少使用，基本没有意义。

2.5.2 IPv6 多播路由协议

IPv6 提供了丰富的多播协议支持，包括 MLDv1、MLDv1 Snooping、PIM-SM、PIM-DM。

1. 多播管理协议

Multicast Listener Discovery for IPv6（简称 MLD）为 IPv6 多播监听发现协议。多播 IPv6 路由器采用 MLD 来发现与其直连的 IPv6 多播监听者，并进行组成员关系的收集和维护。MLD 将收集的信息提供给 IPv6 路由器，使多播分组传送到存在 IPv6 监听者的所有链路上。

MLD 协议源自 IGMP 协议，协议行为完全相同，但与 IGMP 有不同之处，即：IGMP 协议采用 IP 协议号为 2 的报文类型，而 MLD 协议采用 ICMPv6（IP 协议号为 58）的报文类型；IGMP 协议报文使用 IPv4 地址，而 MLDv1 的协议报文地址使用 IPv6 地址；IGMP 的离开报文是 IGMP Leave，而 MLDv1 的离开报文是 Multicast Listener Done。

MLD Snooping 与 IPv4 的 IGMP Snooping 基本相同，唯一的区别在于协议报文地址使用 IPv6 地址。

2. 多播路由协议

IPv6 PIM 协议与 IPv4 PIM 协议的协议行为基本相同，运用潜在的单播路由为多播树的建立提供反向路径信息，并不依赖于特定的单播路由协议。IPv6 PIM 也支持 SM、DM 和 SSM 这 3 种模式。IPv6 PIM 协议的特别之处在于：IPv6 PIM 发送链路本地范围和全球范围的协议报文时，报文的源 IPv6 地址分别使用发送接口的链路本地地址和全球单播地址。

IPv6 多播不支持 MSDP 协议，对于域间 IPv6 多播路由信息的传递，可以使用 IPv6 的 MBGP 协议，其与 IPv4 的 MBGP 协议基本相同。

2.6　IPv6 接入技术

在 IPv4 宽带接入网中，地址分配有无状态地址分配和有状态地址配置（DHCP）两种方式。前者本文已有描述，本章将探讨有状态地址分配方式。认证方式主要分为 PPPoE 和 IPoE 两种。

2.6.1　DHCPv6

DHCPv6（Dynamic Host Configuration Protocol for IPv6）实现为主机动态分配 IPv6 地址和其他网络配置参数的目的。根据客户端需求，自动为客户端配置地址、前缀（DHCP-PD 机制）或者 DNS/NTP 服务器地址等参数。这里有几个常用的概念。IA（Identity Association）：一组分配给 DHCPv6 主机客户端的地址集合。IA 又分为 IA-NA（Identity Association-Network Address）地址集合和 IA-PD（Identity Association-Prefix Delegation）前缀集合。

1. 基本知识

DHCPv6 使用 UDP 消息，DHCPv6 客户端在 UDP 端口 546 上侦听 DHCP 消息，DHCPv6 服务器和中继代理在 UDP 端口 547 上侦听 DHCPv6 消息。而 DHCPv6 中采用多播地址 FF02::1:2 来表示链路范围内所有的 DHCPv6 服务器和中继。

DHCPv6 的消息包含 Solicit、Advertise、Request、Confirm、Renew、Rebind、Reply、Release、Decline、Reconfig、Information-request、Relay-forw 和 Relay-repl 等 13 种，各消息的详细用途将在下面讨论。报文格式有两种，在客户端和服务器之间使用的消息格式如图 2-35 所示。

① msg-type（消息类型）：表明消息的类型。

② transaction-id（事务 ID）：由客户端确定，并用于对 DHCPv6 消息及其交换的消息进行分组。

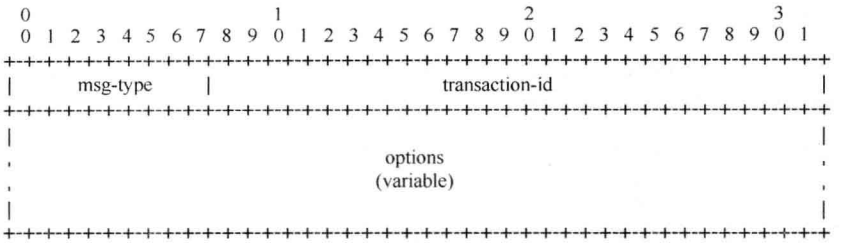


图 2-35　DHCPv6 报文格式

③ Options（选项）：用于指明客户端和服务器的标识、地址以及其他配置设置。

中继和服务器之间使用的报文格式如图 2-36 所示。

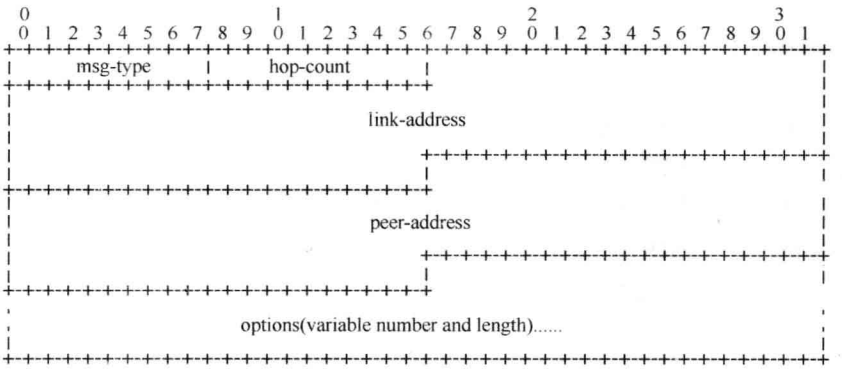


图 2-36　中继报文格式

① msg-type（消息类型）：表明消息的类型。

② hop-count（中继个数）：指明已接收消息的中继代理数。如果其超过了配置的最大数值，中继代理可以放弃该消息。

③ link-address（链路地址）：该字段包含分配给连接到客户端所在子网的接口的非链接本地地址，根据“链路地址”字段，服务器可以确定为客户端合适的地址范围。

④ peer-address（中继地址）：该字段包含最初发送消息的客户端或之前中继该消息的中继代理的 IPv6 地址。

⑤ options（选项）：包含将被中继的消息和其他选项，提供将在客户端和服务器之间进行交换的消息的封装。

2．工作机制

DHCP 实现有状态地址配置和其他信息的配置（路由器公告中的 M 和 O 标记均设置为“1”），其工作机制如图 2-37 所示。

终端客户多播发送 DHCPv6 Solicit 报文，携带客户端 DHCP 唯一标识选项（DUID）和用于配置客户端的选项等选项信息，并定位 DHCPv6 服务器。DHCP 服务器收到 Solicit 报文后，添加服务器 DHCP 唯一标识选项（DUID）和优先级选项，响应 Advertise 消息，表明自身可用性。客户端根据 DHCP 服务器优先级和能提供的服务，选择一台 DHCP 服务器。多播发送 Request 报文，该报文携带客户端 DUID、选定服务器的 DUID 和相关配置选项。拥有相同 DUID 的服务器回送 Reply 报文，携带配置选项。客户端根据 Reply 报文中的选项信息完成自动配置。

DHCPv6 客户端通过链路范围内的多播地址与 DHCPv6 服务器通信。如果服务器和客户端不在同一个链路范围内，则需要通过 DHCPv6 中继来转发报文。DHCPv6 中继和 DHCPv6 服务器之间交互两种报文：Relay-forw 和 Relay-repl。DHCPv6 客户端通过 DHCPv6 中继转发报文，获取 IPv6 地址和其他网络配置参数的过程如图 2-38 所示。

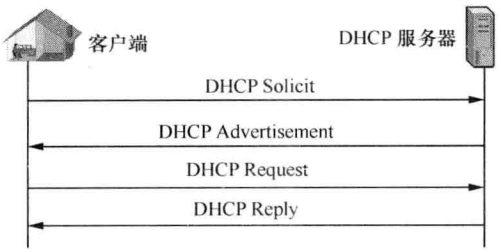


图 2-37　DHCPv6 工作流程图

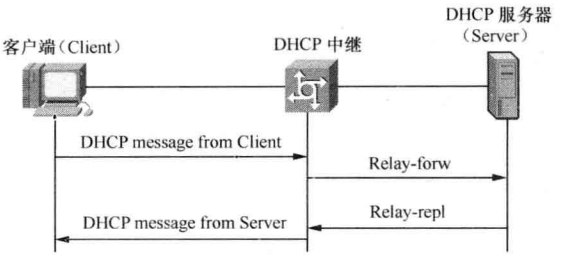


图 2-38　DHCPv6 Relay 工作流程图

DHCPv6 客户端向所有 DHCPv6 服务器和中继的多播地址 FF02::1:2 发送请求。DHCPv6 中继接收到请求后，将其封装在 Relay-forw 报文的中继消息选项（Relay Message Option）中，并将 Relay-forw 报文发送给 DHCPv6 服务器。DHCPv6 服务器从 Relay-forw 报文中解析出客户端的请求，为客户端选取 IPv6 地址和其他参数，构造应答消息，将应答消息封装在 Relay-repl 报文的中继消息选项中，并将 Relay-repl 报文发送给 DHCPv6 中继。DHCPv6 中继从 Relay-repl 报文中解析出服务器的应答，转发给 DHCPv6 客户端。客户端根据 DHCPv6 服务器分配的 IPv6 地址和其他参数进行网络配置。

无状态地址配置能够实现终端地址的自动配置，但其他信息的配置，如 DNS/NTP 等服务器信息的配置（ND 机制新定义了一个 Option25 选项，在 RA 报文携带 DNS 配置信息，实现主机 DNS 配置，但大多数主机不支持该选项），通常还需要借助 DHCP 机制（路由器公告中的 M 设置为“0”，O 设置为“1”），其工作机制如图 2-39 所示。

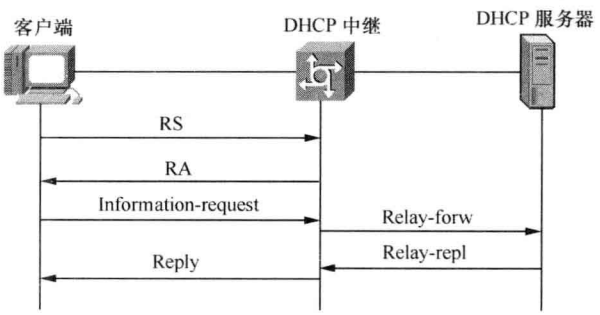


图 2-39　无状态 DHCPv6 工作流程图

客户端通过无状态地址配置方式实现地址配置，即客户端由自身 MAC 根据 EUI-64 生成 64 位接口 ID。链路本地路由器多播发送 RA 报文，并携带前缀选项。客户端根据前缀和接口 ID 形成完整 IPv6 地址。为获得其他配置信息，客户端多播发送 Information-request 报文，请求配置信息。中继路由器将该报文封装到 Relay-forw 报文，发送给 DHCPv6 服务器（通常单播发送）。服务器响应发送携带相应配置信息的 Relay-repl 报文。中继解析出 Reply 报文，并发送给客户端。客户端根据 Reply 报文中提供的参数完成客户端无状态配置（如果接收到多

个 Reply 报文，客户端将选择最先收到的 Reply 报文）。

2.6.2 PPPoE

随着 Internet 和新业务的快速发展，IPv4 逐渐力不从心，IPv6 必然将取代 IPv4。个人宽带业务的需求使 PPPoEv6 成为必然。同时，未来双栈终端大量存在，PPPoE 也要求一次拨号实现同时下发 IPv4 和 IPv6 地址的功能。

1. PPPoEv6

用户 PPPoEv6 接入方式如图 2-40 所示，主要分为 6 个阶段：会话发现、LCP 协商、认证、地址分配、PPPoEv6 通信、Keep-alive 阶段。

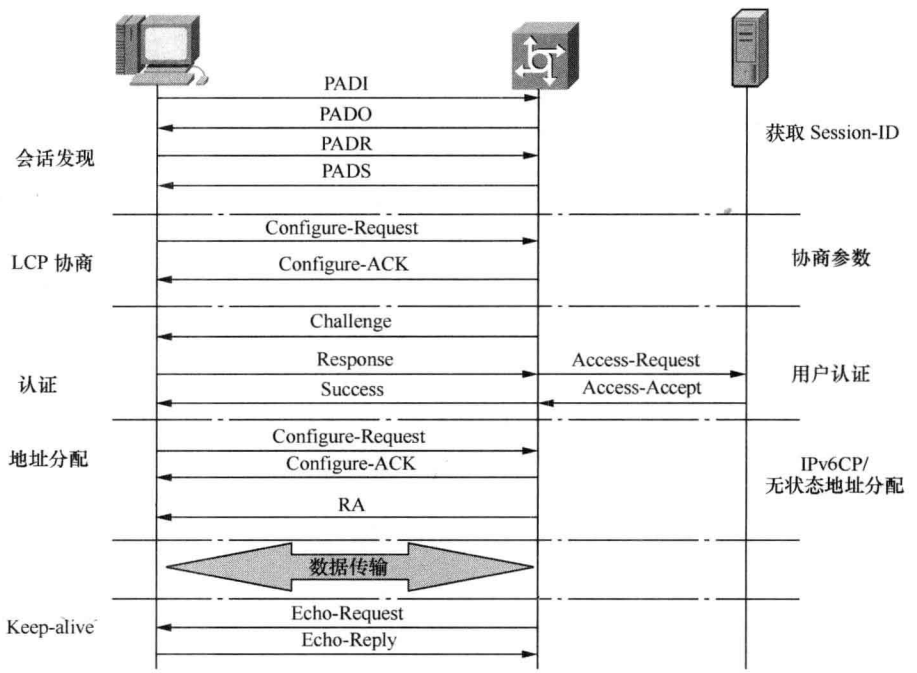


图 2-40 PPPoEv6 过程

① 会话发现阶段：主机和宽带接入服务器之间通过 PADI、PADO、PADR、PADS 报文交换，最终获得 BRAS 分配的唯一 Session-ID。

② LCP 协商阶段：主要协商链路参数，包括最大分组长、认证方式（PAP/CHAP）等。

③ 认证阶段：主机需要和 BRAS、RADIUS 服务器进行信息交互。用户上报用户名和密码信息，Radius 服务器接受用户信息并进行认证。认证通过后，发送 Access-Accept 报文给 BRAS，BRAS 发送认证成功信息给用户。如果认证没有通过，则中断通信。

④ 地址分配阶段：主机通过 IPv6CP 协商获取链路范围内具有唯一性的接口 ID，再通过无状态地址分配方式获取地址前缀。接口 ID 和地址前缀结合形成用户 IPv6 地址。

⑤ Keep-alive 阶段：也称链路检测。BRAS 周期性地下发 Echo-Request 报文，用户在规定时间内进行回应，发送 Echo-Replay 报文，否则 BRAS 认为用户已经下线，停止计费。

地址分配阶段采用无状态地址分配方式配置用户地址，其他配置信息可借助 DHCPv6 机

制实现，即客户端发送 Information-request 报文请求配置信息，DHCP 服务器响应，回复 Reply 报文，为客户端配置 DNS 服务器等信息。地址分配阶段也可以采用有状态地址分配方式，IPv6CP 完成协商接口 ID 后，由 DHCPv6 为客户端分发 IPv6 地址和网络前缀。其工作机制在上节已有描述。客户端与服务器之间交互报文携带前缀相关选项（IA-PD、IA-PD Prefix 等）和 DNS 配置选项，网络前缀加接口 ID 组成客户端 IPv6 地址。

2．PPPoE 同时下发 IPv4/IPv6 地址

PPPoE 同时下发 IPv4/IPv6 地址，有以下两种应用场景。

① 用户终端通过三层家庭网关连接 SP 网络。家庭网关负责与 BRAS 建立 PPP 连接（家庭网关需要支持 IPv4/IPv6 双栈）。

② 用户终端通过二层家庭网关连接 SP 网络。用户终端负责与 BRAS 建立 PPP 连接（家庭网关不感知 IP 层）。

在这两种应用场景下，家庭网络由 IPv4 Only 和双栈 IPv4/IPv6 设备组成。骨干网络提供 IPv4/IPv6 双栈连接。连接家庭网关和 BRAS 的二层接入网络不感知 IP 层。为了满足这两种场景，单一 PPP 连接需要传输 IPv4 和 IPv6 流量。发起 PPP 连接的用户设备（家庭网关或用户终端）需要获得 IPv4 和 IPv6 两个地址，其协议栈如图 2-41 所示。

PPPoE 在 LCP 协商阶段建立 PPP 连接。该连接独立于 PPP 承载的网络层协议。因此，PPP 可以用来同时承载 IPv4 和 IPv6 流量。用户认证通过之后，PPPoE 执行网络层参数配置（地址、DNS 等）阶段，同一 PPP 承载两个不同的网络层协议：IPCP 和 IPv6CP。同一 PPP 连接中，IPCP 和 IPv6CP 并行运行，IPCP 协商分发 IPv4 地址，IPv6CP 协商链路本地地址（接口 ID），SLAAC 或者 DHCP-PD 实现全球唯一 IPv6 地址的配置。

我们以家庭网关发起 PPP 连接为例，介绍一次拨号同时下发 IPv4/IPv6 地址的工作过程。工作流程如图 2-42 所示。

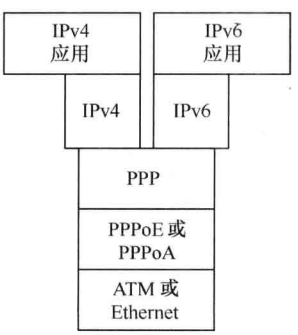


图 2-41　PPPoE 同时下发 IPv4/IPv6 协议栈

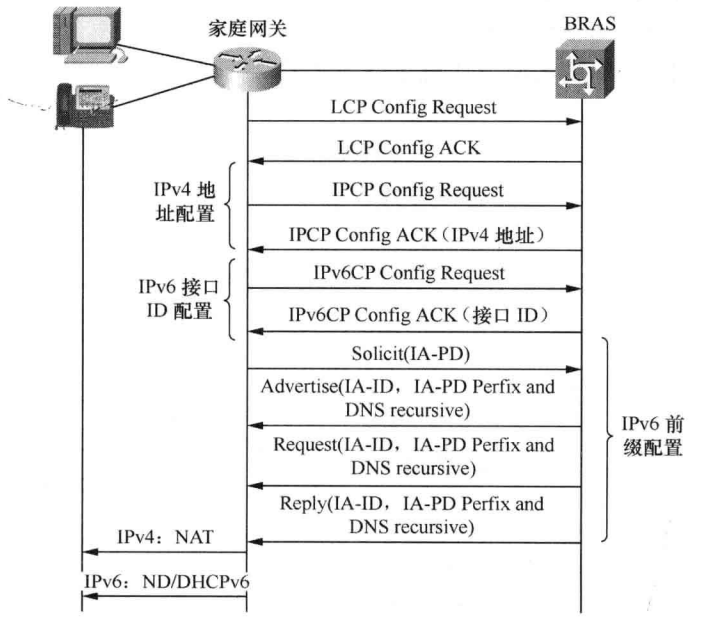


图 2-42　PPPoE 同时下发 IPv4/IPv6 地址过程

PPPoE 会话发现、LCP 协商、用户认证都与传统 PPPoE 过程相同。而地址配置阶段，IPCP 和 IPv6CP 并行运行，IPCP（PPP 头部协议域为 0x0021）分发 IPv4 地址，IPv6CP（PPP 头部协议域为 0x0057）分发链路本地地址（即接口 ID），完整 IPv6 地址的配置借助 SLAAC 或者 DHCP-PD 机制实现（详细机制上文已有介绍）。家庭网关获得完整 IPv4 和 IPv6 地址后，用户 IPv4 终端通过 NAT 机制获得私有 IPv4 地址，用户 IPv6 终端通过 ND 机制（SLAAC）获得地址，借助无状态 DHCPv6 机制获取其他配置信息。双栈用户终端通过 PPPoE 拨号获得 IPv4 和 IPv6 地址，实现机制与上文类似，终端直接通过 PPPoE 获得地址等配置信息，无需再借助 NAT/ND/DHCP 机制。

2.6.3 IPoEv6

IPoEv6 实现方式主要通过电路信息 + 用户名实现。电路信息实现用户的位置绑定，实现方式有两种：通过 DHCPv6 选项信息和通过 QinQ 双层 VLAN 标识。用户名通常使用客户端的 DUID 来标识。通过 QinQ 双层 VLAN 标识用户位置实现技术比较清晰，我们主要描述通过 DHCPv6 选项信息实现用户位置绑定的技术。

IPoEv6 的实现借助 3 个选项：Option1（Client Identifier）、Option18（Interface-ID）和 Option37（Remote-ID）。Option1（Client Identifier）携带客户端 DUID 作为用户名，而 Option18（Interface-ID）和 Option37（Remote-ID）实现用户位置定位。

DUID（DHCP Unique Identifier，DHCP 唯一标识符）是唯一标识一台 DHCPv6 设备（包括客户端和服务器）的标识符。DUID 分为 3 种类型：DUID-LL（DUID Based on Link-layer Address，基于链路层地址的 DUID）、DUID-LLT（DUID Based on Link-layer Address Plus Time，基于链路层地址和时间的 DUID）和 DUID-EN（DUID Based on Enterprise Number，基于企业序列号的 DUID）。DUID-LL 携带的值为设备的 MAC 地址，其实现简单，在 IPoEv6 实现中建议采纳此类型 DUID。

DHCPv6 Option37 和 Option18 选项为 Relay 选项（只能由 DHCP Relay 设备添加），Option37（Remote-ID）用来唯一标识 Relay 设备，Option18（Interface-ID）用来标识 Relay 设备的某一接口。Option37/18 作用类似于 DHCPv4 Option82 的两个子选项 Remote-ID 和 Circuit-ID。

IPoEv6 的实现工作机制如图 2-43 所示。

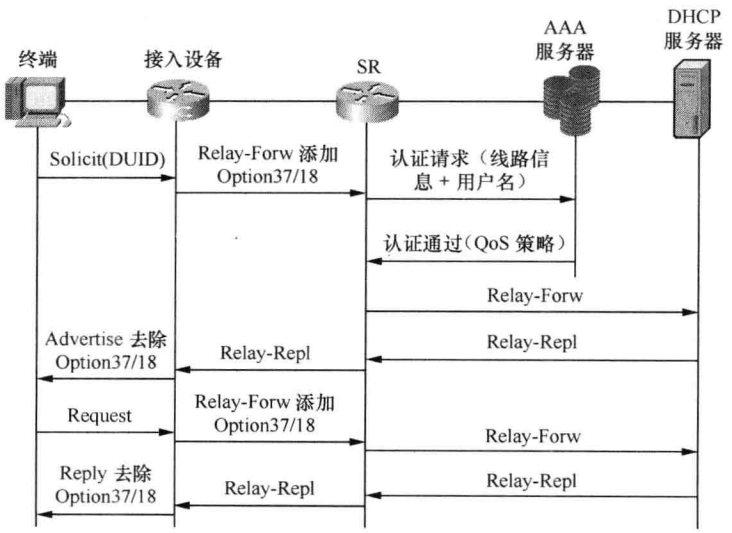


图 2-43 IPoEv6 工作流程图

① 客户端多播发送 Solicit 报文，并携带 DUID、地址配置相关选项和其他选项。

② 接入设备作为 DHCP Relay 设备，将 Solicit 封装在 Relay-forw 报文的中继消息选项（Relay Message Option）中，并添加 Option37/18 选项，实现用户定位，转发中继报文 Relay-Forw 至 SR。

③ SR 提取 Relay-Forw 报文中的选项信息，将 DUID 作为用户名，Option37/18 选项信息作为线路信息，提交 AAA 服务器认证。

④ 认证通过，SR 作为 DHCP Relay，将 Relay-Forw 嵌套封装在“外层” Relay-forw 报文的中继消息选项中，并添加 Option37/18 选项，同时发送中继报文 Relay-Forw 至 DHCPv6 服务器。

⑤ DHCPv6 服务器响应 Advertise 报文表明其可用性。由于收到的 Relay-Forw 为二层封装，需要将 Advertise 报文封装到 Relay-Repl 报文的中继消息选项中，并添加“内层”Option37/18 选项（与接入设备添加选项相同），然后再将 Relay-Repl 封装到“外层” Relay-Repl 报文的中继消息选项中，添加 Option37/18 选项（与 SR 添加选项相同），最后，发送 Relay-Repl 报文至 SR。

⑥ SR 收到响应报文 Relay-Repl 后，剥去“外层” Option37/18 选项，解析出“内层” Relay-Repl 报文，并转发至接入设备。接入设备剥去 Option37/18 选项，解析出 Advertise 报文，并转发至用户终端。

⑦ 客户端根据 DHCPv6 服务器优先级和能够提供的服务，选择一台 DHCPv6 服务器，多播发送 Request 报文，接入设备和 SR 执行与步骤②和步骤④相同的动作。

⑧ DHCPv6 服务器响应 Reply 报文，根据 Option37/18 选项信息、中继报文提供的参考信息等，为客户端分配地址、DNS 服务器地址等信息，然后执行步骤⑤相同的动作，封装 Reply 报文，并发送 Relay-Repl 报文至 SR。SR 和接入设备执行与步骤⑥和步骤⑦相同的动作。最终接入设备转发 Reply 报文至用户终端。

⑨ 用户终端收到 Reply 报文，完成配置。

⑩ IPoEv6 借助 DHCPv6，通过选项实现用户的认证和地址分配。

第 3 章
IPv6 多播技术

3.1 IPv6 多播技术基础

多播路由协议提出的比较早，几乎是与单播路由协议同时提出的，但提出后的一段时间发展比较缓慢，大规模的商用不多。这主要是因为任何协议的商用取决于业务，而多播的业务是视频、音频业务，网络在发展的初期并没有开展这些业务的条件和需求。随着网络的大规模应用，人们对网络的要求越来越多，不仅需要在网络中浏览网页，还需要从网络中收看视频。随着 IPTV 的大规模商用，多播的使用也越来越多。

出现于 IPv4 时代的多播技术，由于其有效解决了单点发送、多点接收的问题，实现了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，因此在 IPv6 中的应用得到了进一步的丰富和加强。

3.1.1 多播的优点

多播是点到多点的传送，单播是点到点的传送，即传统的单播路由必须确定每一个目的地的地址，并分别发送不同的数据分组，而多播则可以通过一个多播地址只发送一个数据分组的拷贝即可。这可以大大节约路由器之间的带宽。

广播是指在一个网络环境中，一对所有的通信方式不仅不安全而且容易引起广播风暴导致网络阻塞。相对于广播而言，多播只对特定的组地址进行数据的传输，解决了广播的缺点。值得说明的是，IPv6 中不再设有广播地址。广播的一部分功能由多播来承担。单播和多播如图 3-1 和图 3-2 所示。

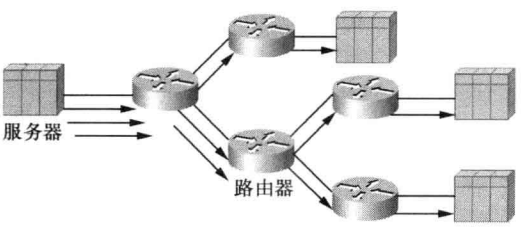


图 3-1 单播示例

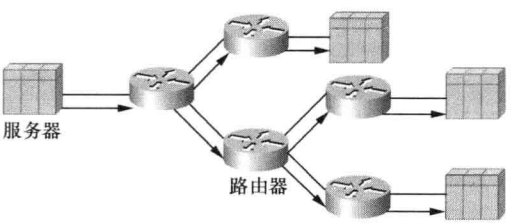


图 3-2 多播示例

3.1.2　多播分发树

多播分发树在 IPv6 多播技术里面是一个重要的概念。多播数据就是通过这个多播分发树来发送多播数据的。多播的分发树有共享树和信源树两种，分别如图 3-3 和图 3-4 所示。

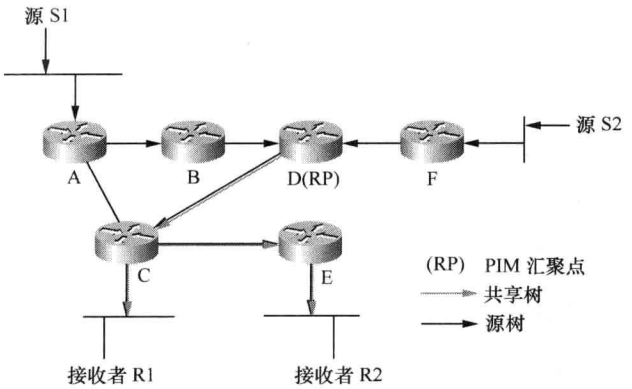


图 3-3　共享树示例

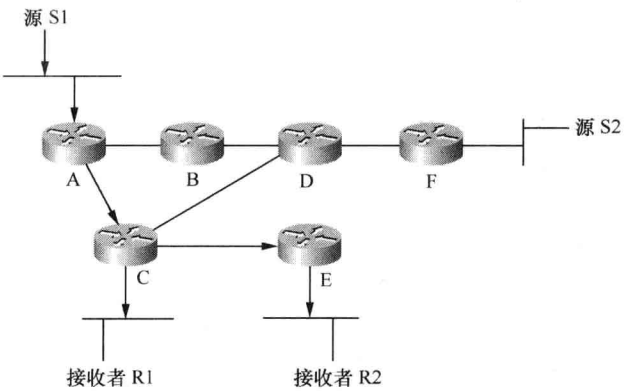


图 3-4　信源树示例

图 3-3 所示为共享树示例，它有以下 3 个特点。

① 以某个路由器作为路由树的树根（该路由器称为汇集点，Rendezvous Point，RP），将 RP 到所有接收者的最短路径结合起来构成的转发树。

② 对应某个组，网络中只有一棵树。

③ 占用内存较少，但路径不是最优的，会引入额外的延迟。

图 3-4 所示为信源树示例，和共享树相比，它也有自己的特点。

① 以多播源作为树根将多播源到每一个接收者的最短路径结合起来构成转发树(从多播源到接收者的最短路径，也称最短路径树 Shortest Path Tree：SPT)。

② 对于某个组，网络为任一个向该组发送报文的多播源建立一棵树。

③ 占用内存较多，但路径最优，延迟最小。

共享树和信源树都有自己的特点，在不同的应用场合有不同的应用，不能一概而论哪种树是最优的。在接收者和源相互不知道的情况之下，可以通过共享树建立分发树，通过共享树可以找到信源树路径；当接收者知道源在哪里的时候，可以直接创建信源树以节约延时和路径。

3.1.3 反向路径转发

单播路由和多播路由是相反的：单播路由关心数据报文要到哪里去，多播路由关心数据报文从哪里来。多播路由使用反向路径转发（Reverse Path Forwarding，RPF）机制转发数据，路由器收到多播数据报文后，只有确认该数据报文是从自己到多播源或者自己到 RP 的出接口上到来的，才转发，否则，丢弃报文。

RPF 具体处理过程如下。

① 在单播路由表中查找到多播报文源的路由。

② 如果该路由的出接口就是报文的入接口，则 RPF 成功。

③ 否则，RPF 检查失败。

3.1.4 多播相关协议

根据协议的作用范围，多播协议分为主机—路由器之间的协议，即多播接收者发现协议，以及路由器—路由器之间的协议，主要是各种路由协议。

由于 IPv6 多播的基本原理和 IPv4 多播并没有区别，因此 IPv6 的组管理协议基本上继承了 IGMP，只是重新起了个名字叫做 MLD（Multicast Listener Discovery）。由于 IGMPv1 实在不堪大用，MLD 只有两个版本。MLD 继承了 IGMPv2，MLDv2 则继承了 IGMPv3。这个继承可以说是全盘继承，只是将报文格式和协议报文的地址进行了细微的调整。由于组管理协议本身并没有本质的改变，相对应的 snooping 机制基本上也是全盘继承过来，改动很小。

PIM 协议可以说是 IP 多播领域的核心协议，其地位不亚于单播路由领域的 OSPF。PIM 在 IPv4 时代经过了充分的发展，协议体系已经非常成熟，IPv6 多播实在是没有理由不用它。不仅仅 PIM-SM 和 PIM-DM 的协议框架被继承了，两个协议中的一些扩展特性如 PIM-SM 的 BSR admin scope、PIM-DM 的状态刷新机制等也都得到了完全的继承。同时 IPv6 中也实现了完整的 SSM 机制。可以说，用于 IPv4 和用于 IPv6 的 PIM，根本就是同一个协议（连协议文本都是合并的）。当然，在报文格式和协议地址等方面还是有区别的。

MSDP 在 IPv4 中被设计用来解决跨域多播的问题，但是，IETF 并没有将 MSDP 迁移到 IPv6 的计划，同时，MSDP 协议文本（RFC 3618）的状态也被置成了 Experimental。MSDP 被废除的原因是多方面的：一是 MSDP 设计为运营商服务，在运营商之间传递多播的（S，G）信息，而由于多播存在的计费监管困难等特殊性，很少有运营商会开通多播服务，MSDP 没有用武之地；二是 SSM 的出现使人们看到了一种能够更简单地实现跨域多播的手段，与 SSM 相比，MSDP 过于复杂了；三是一些新的实现跨域多播的手段的出现，比如将 RP 的地址嵌入多播地址中，等等。在以上几个因素的共同作用下，MSDP 被淘汰了。在 IPv6 中，不再有 MSDP。

3.2 IPv6 多播地址

IPv6 多播地址用来标识一组接口，一般这些接口属于不同的节点。一个节点可能属于 0 到多个多播组。发往多播地址的报文被多播地址标识的所有接口接收。

3.2.1　IPv6 多播地址格式

IPv6 的多播相对于 IPv4 的多播地址有很大的变化。IPv6 地址格式中的第 1 个字节为全“1”的地址标识为多播地址。

IPv6 多播地址格式如图 3-5 所示。

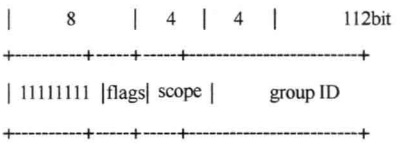


图 3-5　多播地址格式

① flags 是一个 4bit 的字段，可以表示为 0RPT。

最高位为保留位，必须为“0”。

a．R 位取“0”，表示非内嵌 RP 的多播地址；取 1，则表示内嵌 RP 的多播地址，此时 P、T 位也必须置“1”。

b．P 位取“0”，表示非基于单播前缀的多播地址；取 1，则表示基于单播前缀的多播地址，此时 T 位也必须置“1”。

c．T 位取“0”，表示永久分配多播地址；取“1”，则表示非永久分配的多播地址。

② scope：4bit。用来标识此多播组的应用范围，其取值及含义见表 3-1。

表 3-1　　scope 字段的取值及其含义

取值	含义
0、3、F	保留
1	接口本地范围
2	链路本地范围
4	管理本地范围
5	站点本地范围
6、7、9～D	未分配
8	机构本地范围
E	全球范围

③ group ID：112bit，多播组标识号，用来在由 scope 字段所指定的范围内唯一标识多播组，该标识可能是永久分配的，也可能是临时的，这由 flags 字段的 T 位决定。

3.2.2　IPv6 多播地址分配

RFC 3307 建议根据 Group ID 来进行地址分配，见表 3-2。

表 3-2　　多播地址分配建议

Group ID	说明
0x00000001～0x3FFFFFFF	永久组地址，由 IANA 统一分配
0x40000000～0x7FFFFFFF	永久组 ID，特定服务的全球标识符，比如 NTP 使用 0x40404040
0x80000000～0xFFFFFFFF	动态分配组地址

3.2.3　永久分配的 IPv6 多播地址

根据 RFC 4291，目前已被永久分配的 IPv6 多播地址见表 3-3。

表 3-3　　永久分配的 IPv6 多播地址

名称	地址	说明
保留多播地址	FF0x::	不能分配给任何多播组
所有节点多播地址	FF01::1（节点本地） FF02::1（链路本地）	—
所有路由器多播地址	FF01::2（节点本地） FF02::2（链路本地） FF05::2（站点本地）	—
被请求节点多播地址	FF02::1:FFxx:xxxx	由在被请求节点单播或任播地址的低 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

注：以上 x 代表 0～F 的任意一个十六进制数。

3.2.4　基于单播前缀的多播地址

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

基于单播前缀的 IPv6 多播地址的格式如图 3-6 所示，其中各字段的含义如下。

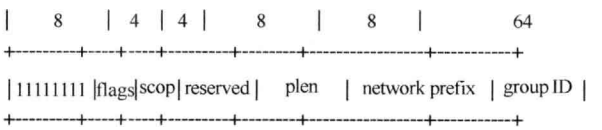


图 3-6　基于单播前缀的 IPv6 多播地址格式

① flags：字段的 R 位置“0”，P、T 位则分别置“1”，表示基于单播前缀的多播地址。

② reserved：8bit。保留字段，必须为 0。

③ plen：8bit。表示网络前缀的有效长度（单位为 bit）。

④ network prefix：64bit。表示该多播地址所属子网的单播前缀，有效长度由 Plen 字段指定。

⑤ group ID：缩短为 32bit，含义不变。

例如：单播前缀为 3FFE:FFFF:1::/48 的网络分配基于单播前缀的多播地址为 FF3x:30:3FFE:FFFF:1::/96（x 表示任意合法的 scope）。

3.2.5　IPv6 SSM 多播地址

IPv6 SSM 多播地址使用基于单播前缀的 IPv6 多播地址格式，其中 plen 和 network prefix 均为 0。因此，IPv6 SSM 多播地址范围为 FF3x::/32（x 表示任意有效的 scope）。

3.2.6　内嵌 RP 地址的多播地址

1．地址格式

内嵌 RP 地址的多播地址也是从基于单播前缀的 IPv6 多播地址演进而来的。其中 flags 字段的 R 位为 1，因此内嵌 RP 地址的多播地址前缀为 FF70::/12。嵌入式 RP（Embedded RP）

是 IPv6 PIM 中特有的 RP 发现机制，该机制使用内嵌 RP 地址的 IPv6 多播地址，使得多播路由器可以直接从该地址中解析出 RP 的地址。

如图 3-7 所示，内嵌 RP 地址的 IPv6 多播地址使用基于单播前缀的 IPv6 多播地址格式，其中各字段的含义如下。

① flags 字段的 R、P 和 T 位均置 1，表示内嵌 RP 地址的多播地址。

② reserved：4bit。保留字段，必须为 0。

③ RIID：4bit。表示 RP 地址的接口 ID。

④ plen：8bit。表示 RP 地址前缀的有效长度（单位为 bit）。

⑤ network prefix：64bit。表示 RP 地址前缀，有效长度由 plen 字段指定。

⑥ group ID：缩短为 32bit，含义不变。

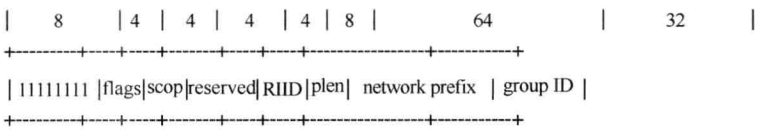


图 3-7　内嵌 RP 地址的 IPv6 多播地址格式

2．计算规则

内嵌于 IPv6 多播地址中的 RP 地址的计算规则如下。

① 先将 IPv6 多播地址 network prefix 字段的前 plen 位作为 RP 地址的网络前缀。

② 再将 IPv6 多播地址 RIID 字段填充到 RP 地址的最低 4bit。

③ 最后，将 RP 地址的所有剩余位补 0。

例如：对于 IPv6 多播地址 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 = 64bit）位，最低 4bit 与 RIID 字段同为 0xF，其余位均为 0，如图 3-8 所示。

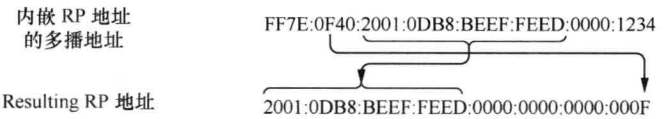


图 3-8　嵌入式 RP 计算举例

3．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 多播地址为 FF7x:y40:2001:DB8:BEEF:FEED::/96，可分配 32bit 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 多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 = 32bit，则此时内嵌 RP 地址的 IPv6 多播地址为 FF7x:y20:2001:DB8::/64，可分配 64bit 的 Group ID，内嵌于其中的 RP 地址为 2001:DB8::y/32。

3.2.7　IPv6 多播 MAC 地址

IPv6 多播 MAC 地址以 0x3333 开头，低 32bit 为 IPv6 多播地址的低 32bit，最终形成 48bit 的多播 MAC 地址。如图 3-9 所示，IPv6 多播地址 FF1E::F30E:101 所对应的多播 MAC 地址

为 33-33-F3-0E-01-01。

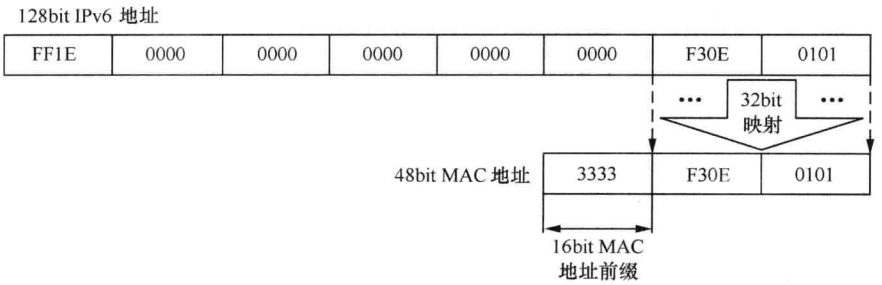


图 3-9　IPv6 多播地址的 MAC 地址映射

3.3　MLD 协议

MLD 是 Multicast Listener Discovery Protocol（多播侦听者发现协议）的简称。它是一个不对称的协议，为多播接收者和多播路由器指定了不同的协议动作，从而实现对多播成员的管理。

3.3.1　MLD 版本

目前 MLD 有以下两个版本：MLDv1 和 MLDv2。MLD 协议源自 IGMP 协议。MLDv1 相当于 IGMPv2 添加了组成员快速离开的机制等；MLDv2 相当于 IGMPv3 增加了可以指定接收或拒绝接收来自某些多播源的报文的成员，以实现对 SSM 模型的支持。

MLD 协议采用 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。

3.3.2　MLD 原理简介

1．MLDv1 加入多播组

① 当主机想接收来自某个组的流量时，它会主动发送 Membership report 报文，为保证加入报文的准确到达，这个 report 会重复发送一两次。

② MLD 查询器收到来自主机的 Membership report 报文，就会创建组成员。

③ 查询器周期性发送普通查询（目的地址为 FF02::1），如果组里还有成员，就会收到主机侧的响应，实现查询响应流程管理组成员，如图 3-10 所示。

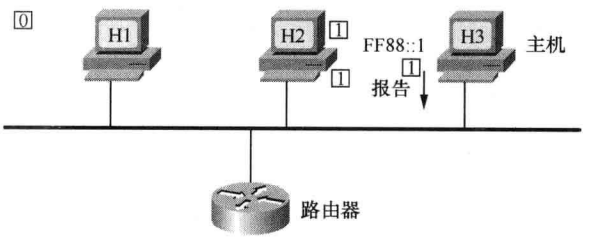


图 3-10　主机加入组

2．主机侧抑制

当一个组中有多个成员时，如图 3-11 中的 H1、H2、H3，在收到一个普通查询报文后，每一个组员启动随机时长的计数器，H2 组员先超时发送响应报文，H1 和 H2 收到 H1 发送的响应报告，就停止自己的计数器，并且不会发送响应报告。结果，子网只会收到一个报文。这个机制称为主机侧 MLD 成员关系报告抑制机制，该机制有助于减少本地网段的信息流量。

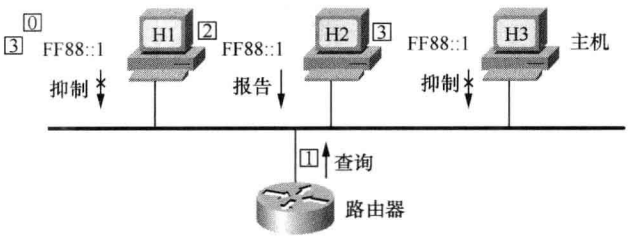


图 3-11　主机侧抑制机制

3．查询者选举

当一个网段内有多台 IPv6 多播路由器时，由于它们都能从主机那里收到 MLD 成员关系报告报文（Multicast Listener Report Message），因此只需要其中一台路由器发送 MLD 查询报文（Query Message）就足够了。这就需要有一个查询器（Querier）的选举机制来确定由哪台路由器作为 MLD 查询器，其选举过程如下。

① 所有 MLD 路由器在初始时都认为自己是查询器，并向本地网段内的所有主机和路由器发送 MLD 普遍组查询（General Query）报文（目的地址为 FF02::1）。

② 本地网段中的其他 MLD 路由器在收到该报文后，将报文的源 IPv6 地址与自己接口的 link-local 地址作比较。通过比较，IPv6 link-local 地址最小的路由器将成为查询器，其他路由器成为非查询器（Non-Querier）。

③ 所有非查询器上都会启动一个定时器（即其他查询器存在时间定时器 Other Querier Present Timer）。在定时器超时前，如果收到了来自查询器的 MLD 查询报文，则重置该定时器；否则，就认为原查询器失效，并发起新的查询器选举过程。

4．离开多播组机制

当一个主机离开某 IPv6 多播组的执行过程如下。

① 该主机向本地网段内的所有 IPv6 多播路由器（目的地址为 FF02::2）发送离开组（Done）报文。

② 当查询器收到该报文后，向该主机所声明要离开的那个 IPv6 多播组发送特定组查询（Multicast-Address-Specific Query）报文（目的地址字段和组地址字段均填充为所要查询的 IPv6 多播组地址）。

③ 如果该网段内还有该 IPv6 多播组的其他成员，则这些成员在收到特定组查询报文后，会在该报文中所设定的最大响应时间（Maximum Response Delay）内发送成员关系报告报文。

④ 如果在最大响应时间内收到了该 IPv6 多播组其他成员发送的成员关系报告报文，查询器就会继续维护该 IPv6 多播组的成员关系；否则，查询器将认为该网段内已无该 IPv6 多播组的成员，于是不再维护这个 IPv6 多播组的成员关系，如图 3-12 所示。

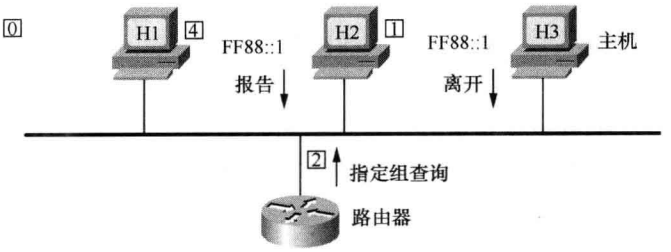


图 3-12　离开多播组

3.3.3　MLDv2 原理简介

在 MLDv1 的基础上，MLDv2 主要增加了 Group-and-Source-Specific-Query，它允许一个组不仅可以用组地址来识别，而且也能用源地址来识别。

1．MLDv2 报文格式

MLDv2 包括查询报文（类型值为 130）和成员关系报文（类型值为 143），MLDv2 没有保留 MLDv1 中的成员离开报文，组成员的离开通过特定格式的成员关系报告实现。

（1）查询报文格式

查询报文格式如图 3-13 所示。

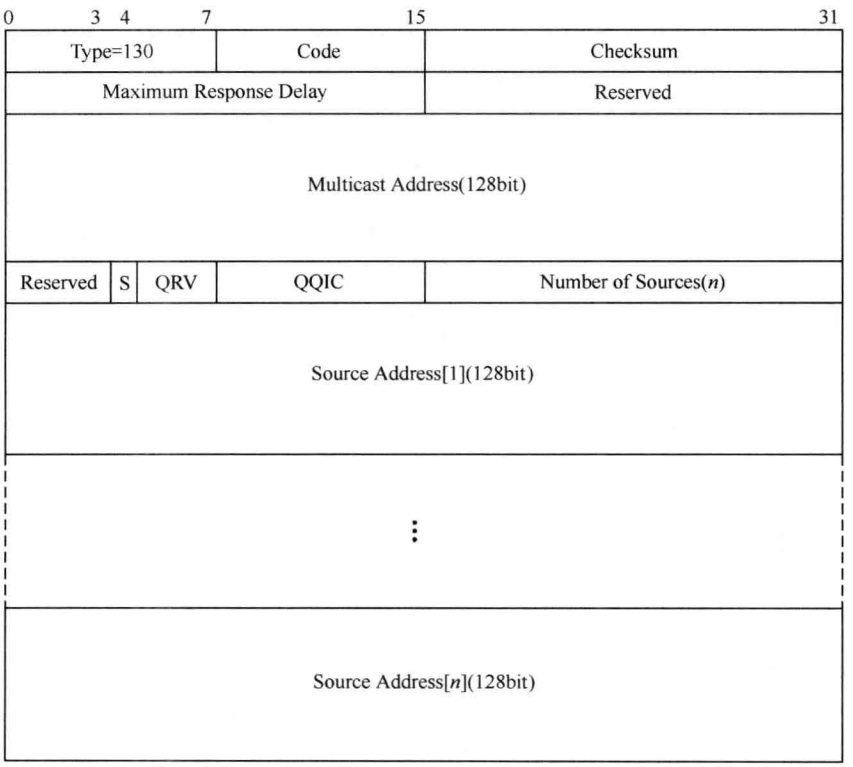


图 3-13　MLDv2 查询报文格式

各字段说明见表 3-4。

表 3-4　　MLDv2 查询报文各字段含义

字段	描述
Type = 130	报文类型，130 代表查询报文
Code	初始化为 0
Checksum	标准的 IPv6 校验和
Maximum Response Delay	主机发送报告报文前允许的最大响应时间
Reserved	保留字段，初始化为 0
Multicast Address	普遍组查询中，此字段设置为 0 特定组或特定源组查询中，此字段设置为待查询的 IPv6 多播组地址
S	标识位，表示路由器接收到查询报文后是否对定时器更新进行抑制
QRV	查询器的健壮性变量（Querier's Robustness Variable）
QQIC	查询器发送普遍组查询报文的查询间隔（Querier's Query Interval Code）
Number of Sources	普遍组查询或特定组查询中，此字段设置为 0 特定源组查询中，此字段表示查询报文中包含的源地址个数
Source Address[i]	特定源组查询中的 IPv6 多播源地址（$i = 1, 2, \cdots, n$，其中 n 表示源地址的个数）

（2）报告报文格式

主机通过发送 MLD 报告报文来汇报当前的多播侦听状态。MLD 报告报文的格式如图 3-14 所示。

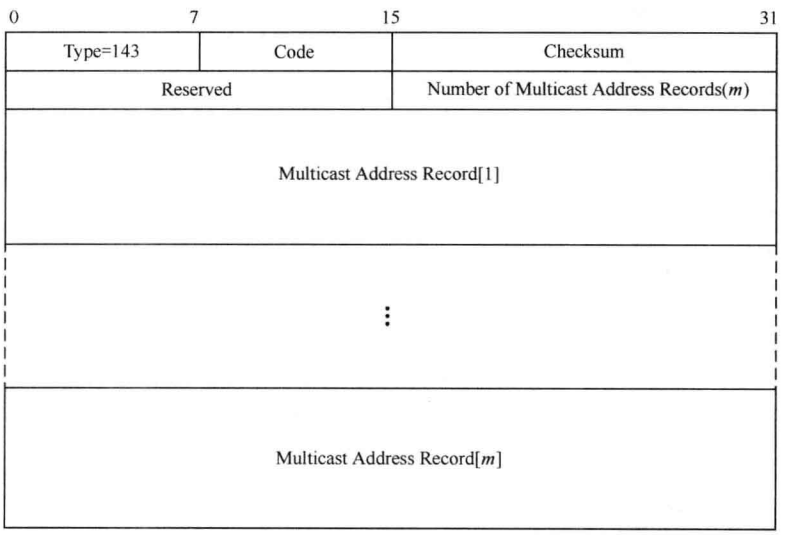


图 3-14　MLDv2 报告报文格式

各字段说明见表 3-5。

表 3-5　　MLDv2 报告报文各字段含义

字段	描述
Type = 143	报文类型，143 代表报告报文
Reserved	保留字段，发送时设置为 0，接收时忽略此值

续表

字段	描述
Checksum	标准的 IPv6 校验和
Number of Multicast Address Records	IPv6 多播地址记录的个数
Multicast Address Record[i]	多播地址记录，表示主机在接口上侦听到的每个 IPv6 多播地址信息，包括记录类型、IPv6 多播地址、IPv6 源地址等（$i = 1, 2, \cdots, m$，其中 m 表示 IPv6 多播地址记录的个数）

2. MLDv2 组模式信息

IPv6 多播路由器为每一个接口上的每一个多播地址（多播组）维护相关的状态信息。这些多播组的状态主要由以下信息构成：（group address，group timer，filter mode，[source record]）。

其中，每一个 source record 的基本构成如下：

（source address，source timer）

MLDv2 中组包含 INCLUDE 和 EXCLUDE 两种模式：INCLUDE 只有来自 source-list 参数中列出的那些 IP 源地址并发往指定多播地址的数据报才会被接收；EXCLUDE 只有来自除 source-list 参数中列出的那些 IP 源地址之外的源地址，并发往指定多播地址的数据才会被接收。如果需要接收某一个多播地址的所有源的数据，则使用一个空 source 列表，并设置其模式为 EXCLUDE，这就是对旧版本的兼容。

考虑到和 MLDv1 的兼容，MLDv1 加入操作等效于：IPMulticastListen（socket，interface，multicast-address，EXCLUDE{}）；

而离开操作等效于：IPMulticastListen（socket，interface，multicast-address，INCLUDE{}）；

这里{}表示一个空的源列表。

group timer 仅在 EXCLUDE 模式使用，表示从 EXCLUDE 切换到 INCLUDE 的时间间隔。每个组地址使用一个。

模式定时器描述以及相关动作如下：

INCLUDE　不使用　所有接收者在 include 模式下

EXCLUDE　timer>0　至少一个接口者在 exclude 模式下

EXCLUDE　timer = 0　该组地址没有接口在 exclude 模式下，如果 request list 为空，删除组地址；如果不为空，切换到 IN 模式，request list 移到 include list，删除 exclude list。

3.4 IPv6 PIM 协议

IPv6 PIM 协议与 IPv4 PIM 协议基本相同，除了报文中的 IP 地址结构不同，IPv6 PIM 协议也支持 PIM-SM、PIM-DM 和 PIM-SSM 这 3 种模式。

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。

3.4.1　PIM 消息格式

PIM 消息格式如图 3-15 所示。

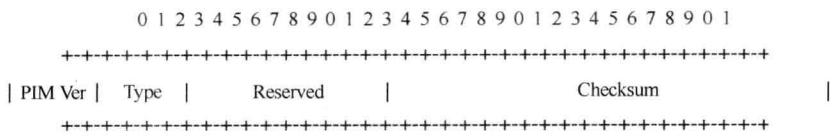


图 3-15　PIM 消息格式

① PIM Ver：PIM 版本号，其值为 2。

② Reserved：保留，发送时置“0”，接收时忽略。

③ Checksum：标准 IP 校验和，校验范围是整个 PIM 消息，但不包括 Register 消息中的多播数据报文。

④ Type：PIM 消息类型，具体见表 3-6。

表 3-6　　PIM-SM 消息类型

值	类型	传播方式
0	Hello	Multicast to ALL-PIM-ROUTER
1	Register	Unicast to RP
2	Register-Stop	Unicast to source of Register packet
3	Join/Prune	Multicast to ALL-PIM-ROUTER
4	Bootstrap	Multicast to ALL-PIM-ROUTER
5	Assert	Multicast to ALL-PIM-ROUTER
6	Graft（PIM-DM only）	Multicast to ALL-PIM-ROUTER
7	Graft-Ack（PIM-DM only）	Unicast to source of Graft packet
8	CRP-Advertisement	Unicast to Domain's BSR

3.4.2　PIM 邻居发现和 DR 选举机制

1. Hello 报文发送

PIM 协议中路由器通过发送 Hello 报文向所有的 PIM 路由器宣告自己的存在，通过 Hello 报文，路由器可以建立路由器邻接关系。

路由器启动后，在随机延时后发送 Hello 报文，此后，在每个 Hello_period 周期内，PIM 路由器向所有配置了 PIM 协议的邻居接口周期发送 PIM Hello 报文。

路由器要向未发送过 Hello 报文的 IP 地址发送 Join/Prune 或 Assert 报文时，先立即发送一个 Hello 报文，然后再发送需要发送的报文。

当路由器收到一个新邻居或者一个原先邻居的带有一个新的 GENID 的 Hello 报文时，在随机延时后发送 Hello 报文，以便尽快建立邻居关系。

当接口关闭或者更改 IP 地址时，路由器立即发送 Holdtime 为 0 的 Hello 报文来终止邻居关系。

Hello 报文发送是路由器用于维护邻居关系的一种机制，PIM 注册的每一个接口都要周期性的发送 Hello 报文到所有 PIM 路由器的组地址（224.0.0.13）。同时每个 PIM 接口还要

接收来自邻居接口的 Hello 报文。每个路由器还需要记录来自邻居路由器的 Hello 报文，用于获取邻居路由器信息，也用于 DR 选举，以及协商其他附加功能。Hello 报文格式如图 3-16 所示。

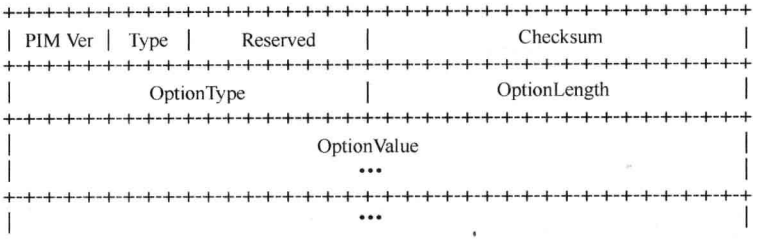


图 3-16　Hello 报文格式

Hello 报文格式中的选项类型（Option Type）主要有：1——Hold Time；2——LAN Prune Delay；19——DR Priority，仅用于 PIM-SM；20——Generation ID；21——状态更新标志。

① Hold Time 用于指示在没有进一步收到 PIM Hello 报文的情况下，接收方何时终止与发送方的相关邻接。邻接关系持续时长通常是发送方发送 PIM Hello 报文周期的 3.5 倍。其格式如图 3-17 所示。

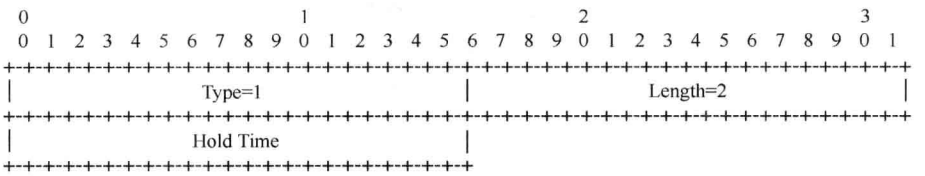


图 3-17　Hold Time 选项格式

Hold Time 值为 0xffff 时，表示接口永久存在，定时器永不会超时。

Hold Time 值为 0x0000 时，表示接口马上 Down 或马上更新 IP 地址。

② LAN Prune Delay 用于调整多路访问网络中的剪枝传输延时。其格式如图 3-18 所示。

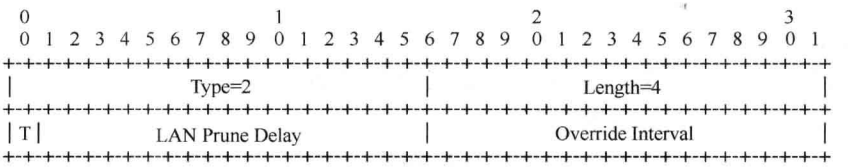


图 3-18　LAN Prune Delay 选项格式

其中 T 用于 PIM-SM，对于 PIM-DM 设置为 0。

③ Generation ID 的格式如图 3-19 所示。

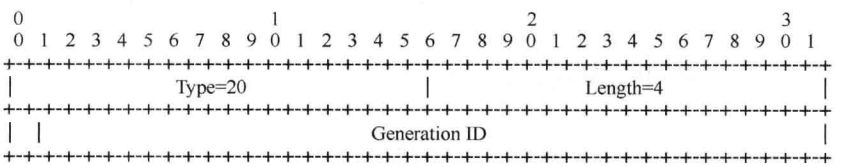


图 3-19　Generation ID 选项格式

④ 状态更新的格式如图 3-20 所示。

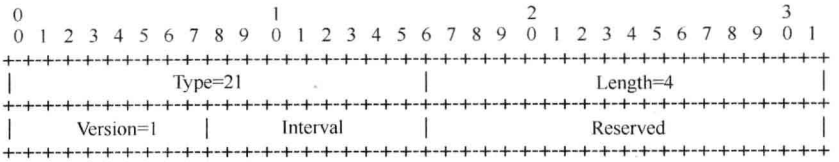


图 3-20　状态更新选项格式

每个 Hello 报文必须包括 DR 优先级，即使接口没有配置 DR 优先级选项，它在发送 Hello 报文的时候都会携带一个默认的 DR 优先级选项，默认的 DR 优先级是 1。

一个 LAN 的剪枝延迟选项应该被包含在 Hello 中，发送到其他的多播接口，通知其他路由器使用设定的值而不是默认的传播延迟。

Generation ID 选项也会包含在 Hello 报文中。每个使用 PIM 协议的接口在重启的时候都会随机生成一个 Generation ID 的 32 位的随机数。当一个携带新的 Generation ID 选项的 Hello 报文到达时，任何旧的 Generation ID 都会被替换，这有可能会导致一个新的 DR 路由器的选举。

直连源的 PIM-DM 路由器周期性地发送状态更新消息，用于刷新其他路由器的状态。

每个 PIM 接口都有一个 Hello 定时器用来触发 Hello 报文的发送。当使用 PIM 协议的路由器接口第一次启动或者重新 Up，为了能使新起来的接口快速地学习到 PIM 邻居的信息，Hello 定时器被设置成 0 到 Triggered_Hello_Delay 之间的一个随机延时，这主要是为了避免大量路由器同时启动导致 Hello 报文风暴。

当一个路由器收到一个新的邻居的 Hello 报文，或者从一个已经存在的路由器收到一个新的 Generation ID 的 Hello 报文时，这个路由器就随机延迟一个 0 到 Triggered_Hello_Delay 的时间再发送 Hello 报文，而不用等到定时器超时。这个过程不用复位定时器。

当一个接口将要被 Down 掉或者更换地址的时候，路由器就立刻发送一个保持时间为 0 的 Hello 报文。如果是更换地址，它就用旧的地址发送这个 Hello 报文，这就会促使邻居将旧的 IP 地址删除。当路由器更换完地址之后，它就必须用新的 IP 主地址立刻发送一个 Hello 报文。

路由器在收到某个路由器的 Hello 报文之前是不会处理这个路由器发送的加入或剪枝消息和断言消息的。如果一个路由器马上要发送加入或剪枝消息，但是它还没向上游路由器发送 Hello 报文，就应立刻发送 Hello 报文并接着发送加入或剪枝消息，而不需要等到 Hello 定时器超时后再发送。

如果路由器要向一个新的邻居发送加入或剪枝报文，或者要对一个新邻居发来的 Assert 报文进行回应的时候，若这个随机的定时器还没有超时，这时就立刻发送 Hello 报文，而不需要等到这个随机定时器超时。

2. DR 选举

当接口收到一个 Hello 报文的时候，会记录以下信息：邻居接口、邻居主地址、邻居的 Generation ID、DR 优先级、DR 优先级标志、超时定时器。

PIM 路由器通过发送 Hello 报文来进行 DR 选举。选举的获胜者成为本网段的 DR。在 PIMv2 的 Hello 报文中 DR 优先级用于 PIM DR 的选取。通过对路由器配置可以指定局域网

上每个路由器的 DR 优先级（缺省为 1），优先级最大的路由器将成为本网段的 DR。如果所有路由器优先级相同，则 IP 地址最大的胜出。如果网络中有路由器不支持 DR 优先级，则仍采用 IP 比较的方式。

收到新的 Hello 报文，邻居超时，路由器的 DR 优先级变化时都会引起接口当前 DR 的选项发生变化。一个路由器成为 DR 或者停止成为 DR，都会引起 DR 注册状态机状态的改变。

3.4.3 PIM-SM 注册

由于 PIM-SM 使用的是信息只能下行的单向共享树，因此要使信息到达 RP，第一跳路由器要通过注册过程通知 RP 多播信息源的存在。RP 处理注册消息，然后把 SPT 与源连接起来，从而使多播信息流送到 RP，然后沿共享树向下转发。

注册过程根据 RP 上是否有接收者，分为接收者先加入和源先注册两种情况。以源先注册为例，解释注册过程如下。

① 第一跳 DR 收到源发送的多播数据分组。

② 第一跳 DR 将多播数据分组封装在 Register 报文中，单播发送到 RP。

③ RP 收到 DR 发送的 Register 报文，由于没有下游接收者，则丢弃分组。并向 DR 单播回送 Register-Stop 报文。

④ 第一跳 DR 停止封装多播数据分组到 Register 报文中，继续向 RP 单播发送空的注册报文，并丢弃从源发送的数据分组。

⑤ 当 RP 上有了下游接收者时，向源发送（S,G）加入消息。

⑥ 直到从源到 RP 之间的 SPT 树创建完成，RP 接收到沿 SPT 向下的（S,G）信息，RP 向源发送 Register-stop 报文，注册过程完成。

第一跳 DR 的状态机如图 3-21 所示。开始注册时，DR 注册状态机处于 Join 状态；当注册完成时，处于 JP 和 Prune 两个状态之间，由定时器和 Register-stop 消息触发状态转换。如果 RP 发生改变，则重新进入 Join 状态，开始新的注册过程。

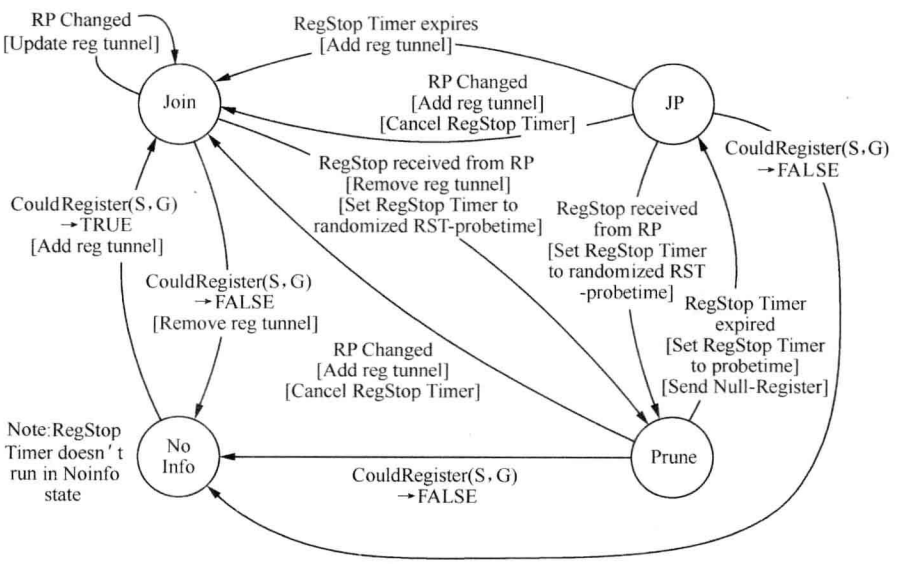


图 3-21 DR 注册状态机

3.4.4　PIM-SM 加入/剪枝

一个路由器要想接收多播数据报文，首先必须要加入这个组。路由器通过发送显式 Join/Prune 消息来加入多播分发树。同样，一个主机或者路由器不需要这个多播数据的时候就发送离开报文。当路由器发现下面没有接收者的时候就会向上发送剪枝报文。在具体的实现中，加入和剪枝报文不是两个独立的报文，而是报文里面的不同字段包含的组和源信息的列表。路由器根据这个列表决定需要加入哪些组和剪枝哪些组。每个 PIM 的 Join/Prune 消息既包含加入列表又包含剪枝列表，其中之一可为空。每个列表都有源地址集，指示出路由器要加入或剪枝的 SPT/RPT，其具体的消息格式如图 3-22 所示。

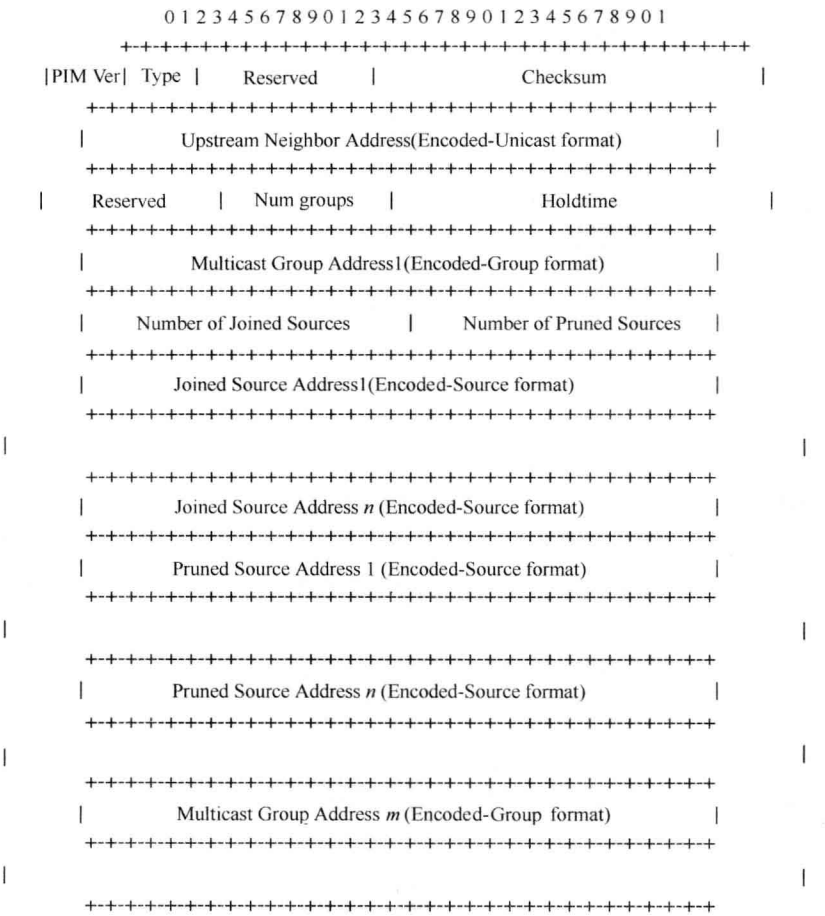


图 3-22　加入/剪枝消息格式

① Upstream Neighbor Address：上游邻居地址。
② Num groups：需要处理的组数目。
③ Holdtime：保持时间，用于周期性的发送加入剪枝报文。
④ Number of Joined Sources：某一个组的加入源个数。
⑤ Number of Pruned Sources：某一个组的剪枝源的个数。
⑥ Joined Source Address 和 Pruned Source Address 是具体的源地址信息。

加入/剪枝有 SPT 树方向上的加入/剪枝和 RPT 方向上的加入/剪枝两种。其区别只是目标不同：SPT 方向上的是朝向源 S 的，而 RPT 方向上的是朝向 RP 节点的。这在报文的上游邻居的填写上面可以看得出区别。相应的，发送的报文就有（S，G）加入/剪枝报文和（*，G）加入/剪枝报文。但是，不管是哪一种加入/剪枝报文，其处理机制都是相同的。

以 RPT 加入为例，如图 3-23 所示，加入过程如下。

① 最后一跳路由器 R3 收到主机 2 发出的 IGMP 加入消息后，向 RP 方向发送（*，G）加入消息。

② 沿途路由器收到加入消息后建立（*，G）条目，并向 RP 方向继续发送加入消息。

③ 这个过程直到 RP 或者沿途路由器已经有了该（*，G）条目时结束。这时从 RP 到最后一跳路由器 R3 的 RPT 建立完成。

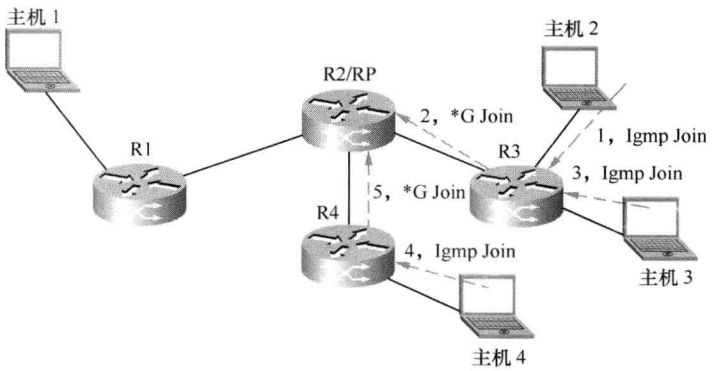


图 3-23　RPT 加入

以 RPT 剪枝为例，如图 3-24 所示，剪枝过程如下。

① 主机 2 和主机 3 都发送了 IGMP 离开消息，最后一跳路由器 R3 发出指定组查询，没有任何主机回应，则 R3 删除（*，G）条目的所有出接口，同时向 RP 方向发送（*，G）剪枝消息。

② 沿途路由器收到剪枝消息，等待收到消息的接口上是否有其他下游路由器发送加入消息。如果没有，则删除（*,G）条目的出接口。如果此时该条目出接口列表为空，则继续向上游发送剪枝消息。

③ 这一过程持续至 RP 或者沿途路由器的（*,G）条目出接口列表不为空为止。

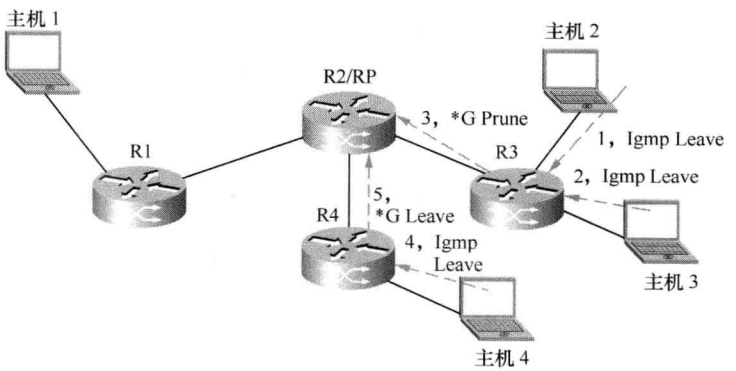


图 3-24　RPT 剪枝

3.4.5　PIM-SM 最短路径树切换

在 PIM-SM 中，最后一跳路由器加入朝向 RP 的共享树，一旦源发送的多播报文到达最后一跳路由器，最后一跳路由器就可以进行向最短路径树切换的操作。决定切换的条件如下：

```
Void
CheckSwitchToSPT(S,G) {
    If  (  (  pim_include(*,G)   (-)  pim_exclude(S,G)
        (+)  pim_include(S,G) ≠NULL  )
            AND SwitchToSPTDesired(S,G) ) {
    # Note: Restarting the KAT will result in the SPT switch
            restart KeepaliveTimer(S,G);
    }
}
```

从上面的伪代码中可以看到，AND 连接的切换的 3 个条件如下。

① 最后一跳路由器的本地成员信息不为空，并且计算的都是和主机直连的接口。pim_include 和 pim_exclude，表示 SPT 切换是从最后一跳 DR 路由器发起的。

② SwitchToSPTDesired（S,G）宏为真。这个宏主要起到决定怎么切换的作用。是永不切换，还是收到数据分组立即切换，还是流量超过一定值切换，可以通过这个宏来调节。

③ 切换是通过 KAT 的重启来实现的，重启 KAT 将直接导致 SPT 的切换。

PIM-SM 最短路径树切换过程如图 3-25 所示。按照 RPT 的转发，主机 4 收到多播数据的途径是 R1—R2—R4。显然这一路径不是最优的。为了实现数据转发路径是 R1—R4，需要进行 SPT 切换。切换过程如下。

a．R4 定期统计多播数据流速率，如果超过某设定好的阈值，那么进行 SPT 切换。

b．R4 上创建（S,G），并向源方向发送（S,G）加入消息。

c．R1 收到加入消息后，数据分别向 R2 和 R4 两个接口转发。

d．R4 收到从 R1 转发的多播数据，证明从 R1 到 R4 的 SPT 树建立完成，则向 RP 发送 RP 位剪枝消息。

e．RP 创建（S,G）条目，出接口只有 R3，这样多播数据就不会从 RP 向 R4 转发，而是沿着 R1—R4 向主机 4 转发了。

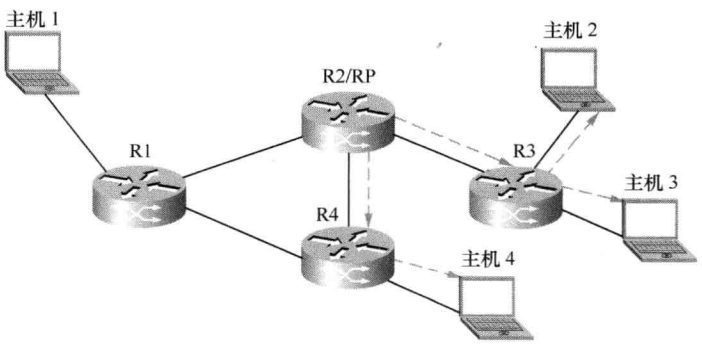


图 3-25　SPT 切换

3.4.6 PIM-DM 加入和剪枝过程

一个路由器要想接收多播数据报文，首先必须要加入这个组。同样，当一个主机或者路由器不需要这个多播数据的时候就发送离开报文，当路由器发现下面没有接收者的时候就会向上发送剪枝报文。在具体的实现中，加入和剪枝报文不是两个独立的报文，而是报文里面的不同字段包含的组和源信息的列表。路由器根据这个列表决定需要加入哪些组和剪枝哪些组。每个 PIM 的 Join/Prune 消息既包含加入列表又包含剪枝列表，其中之一可为空。每个列表都有源地址集，指示出路由器要加入或剪枝的 SPT/RPT，其具体的消息格式如图 3-26 所示。

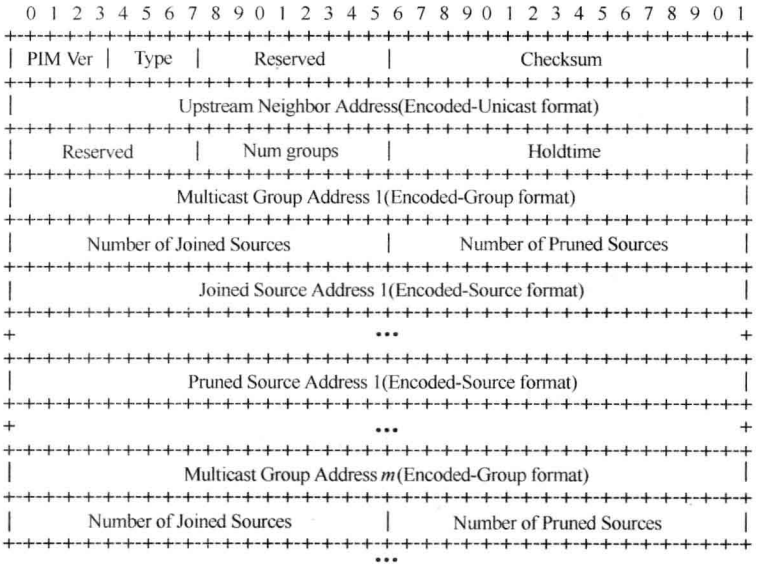


图 3-26 Join/Prune 消息格式

Upstream Neighbor Address：是上游邻居地址。如果是 XG 的报文的话就是 RP 方向上的上游邻居地址；如果是 SG 的 J/P 报文的话就是源 S 方向上的上游邻居地址。

Num groups：需要处理的组个数。

Holdtime：保持时间，用于周期性的发送加入/剪枝报文。

Number of Joined Sources：某一个组的加入源个数。

Number of Pruned Sources：某一个组的剪枝的源的个数。

Joined Source Address 和 Pruned Source Address：具体的源地址信息。

这个组是由紧挨着它的上一个组地址决定。

加入/剪枝有 SPT 树方向上的加入/剪枝和 RPT 方向上的加入/剪枝之分，它们只是目标不同。SPT 方向上的加入/剪枝是朝向源 S 的，而 RPT 方向上的加入/剪枝是朝向 RP 节点的。这在报文上游邻居的填写上面可以看得出区别，相应的，发送的报文就有（S,G）加入/剪枝报文和（X,G）加入/剪枝报文。但是，不管是哪一种加入/剪枝报文，其处理机制都是相同的。

1. PIM-DM 加入

以 SPT 加入为例，如图 3-27 所示，加入过程如下。

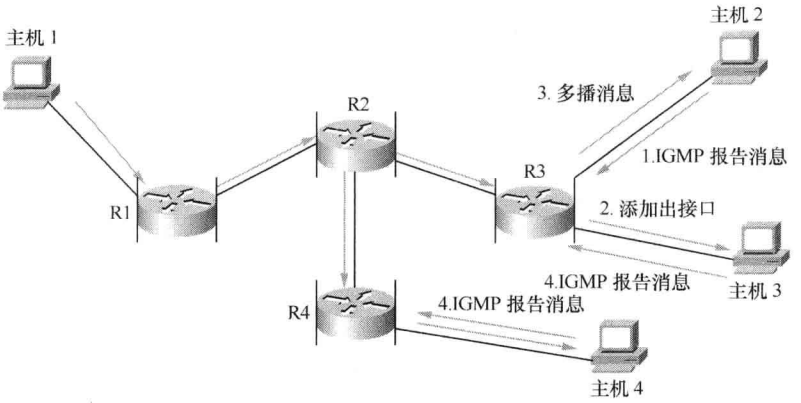


图 3-27　SPT 加入

① 主机 2 首先向网络发送 IGMP 加入消息，申请加入多播组 G。

② 最后一跳路由器 R3 收到主机 2 发出的 IGMP 加入消息后，检查是否有（S,G）条目，如果没有，就不处理；如果存在该（S,G）条目且接收接口是出接口，再检查该出接口是否已经存在，如果没有则添加到对应的（S,G）出接口列表中。重设 IGMP 过期时间，并给该出接口打上直连成员标志。

③ 当源 S 发送多播信息时，路由器 R3 就从对应的出接口向主机 2 转发这些多播信息。

④ 主机 3 和主机 4 加入多播组 G 的过程与主机 2 的加入过程类似。

2．PIM-DM 剪枝

以 SPT 剪枝为例，如图 3-28 所示，剪枝过程如下。

① 主机 2 发送了 IGMP 离开消息后，最后一跳路由器 R3 发出指定组查询。若没有任何主机回应，则 R3 删除该（S,G）条目对应的出接口。

② 主机 3 发送 IGMP 离开消息，最后一跳路由器 R3 发出指定组查询。若没有任何主机回应，则 R3 删除该（S,G）条目对应的出接口。此时 R3 的出接口列表为空，向源 S 方向发送（S,G）剪枝消息。

③ 沿途路由器 R2 收到剪枝消息进入待剪枝状态后，启动待剪枝定时器，等待其他下游路由器向收到消息的接口发送加入消息用于剪枝否决。若 R2 没有收到加入消息，则删除（S,G）条目的出接口。但此时该条目出接口列表为非空，则不继续向上游发送剪枝消息。

④ 主机 4 发送 IGMP 离开消息，最后一跳路由器 R4 发出指定组查询。若没有任何主机回应，则 R4 删除该（S,G）条目对应的出接口。

⑤ 路由器 R4 出接口列表为空，向源 S 方向发送（S,G）剪枝消息。

⑥ R2 收到剪枝消息后进入待剪枝状态，启动待剪枝定时器，等待下游路由器发送加入消息。若定时器超时，而 R2 没有收到加入消息，则删除（S,G）条目的出接口。此时该条目出接口列表为空，则继续向上游发送剪枝消息。

⑦ 路由器 R1 收到剪枝消息后剪枝出接口，R1 的出接口列表为空，但它是直连源，不向源继续发送剪枝消息，而是将收到的源发送的消息做简单的丢弃。

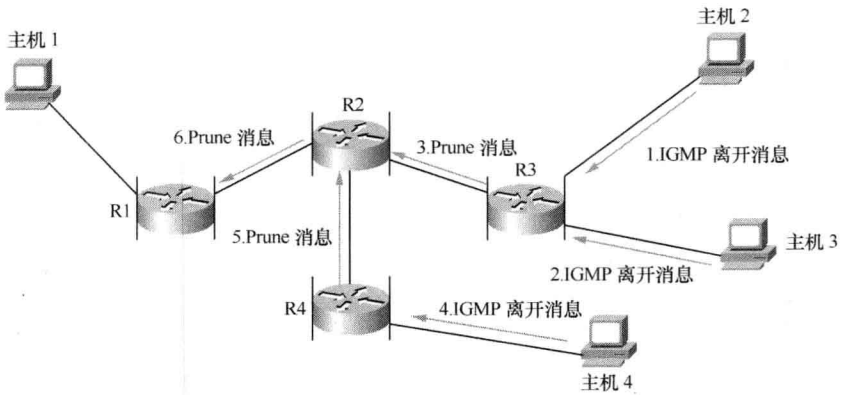


图 3-28　SPT 剪枝

3.4.7　PIM-DM 嫁接机制

如图 3-29 所示，嫁接过程如下。

① 主机 2 发送 IGMP 加入消息。

② R3 收到 IGMP 加入消息，添加出接口。此时出接口列表变为非空，向源方向发送嫁接消息，进入等待嫁接确认状态，并启动 GRT 定时器（GraftRetry Timer）。如果 GRT 超时，那么就重传嫁接消息。

③ 路由器 R2 收到嫁接消息，添加出接口。如果该接口原来是剪枝状态，则取消剪枝状态，并向 R3 发送嫁接确认消息来应答。如果该路由器的出接口列表原来为空，则继续向源方向发送嫁接消息，同时向 R3 发送嫁接确认消息来应答。

④ 路由器 R3 收到嫁接确认应答后，终止 GRT，实现将先前剪枝的分支嫁接回分发树。重新启动路由器 R2 到路由器 R3 间的多播信息流的转发。

⑤ 主机 4 发送 IGMP 加入消息也会引发路由器 R4 向上游发送嫁接消息，具体过程与前面类似。

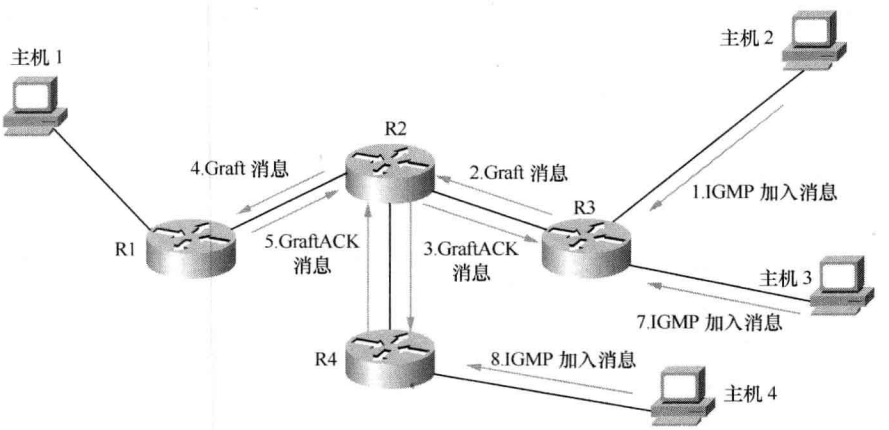


图 3-29　SPT 嫁接

3.4.8　PIM 的断言机制

在同一个子网中有多个 PIM 路由器。这有可能导致一个路由器会有多个上游的情况，在

接收数据的时候会收到多个数据的副本。PIM 本身没有辨别数据分组重复的能力。如果这种情况发生了就会启动断言机制来选取一个上游，避免发送多个数据副本。这种机制是通过 PIM 的 Assert 消息来实现的，同样，这个 Assert 消息也会被下游路由器收到，下游路由器就会立刻向上游路由器发送 Join/Prune 消息来回应这个 Assert 消息。

通常路由器只会接收一个已知邻居的 PIM 的 Assert 消息。如果它收到一个未知 IP 的 Assert 消息，而这个 IP 的路由器没有向路由器发送 Hello 分组，那么这个 Assert 消息就会被丢弃不做任何处理。另外，如果这个邻居路由器的 Assert 被配置成采用 IPSec 验证功能，那么从这个邻居收到的 Assert 消息也必须采用验证功能。

Assert 的具体过程如图 3-30 所示。

① 多播数据分组通过路由器 A 和 B 发送到同一网络中。因此，路由器 A 和 B 分别从它们的出接口接收到多播数据分组，触发它们向网络上发送 Assert 消息。

② 路由器接收到 Assert 消息，根据 assert_metric 选出 Assert Winner。

③ Assert Winner：保留该接口，并负责该网络中转发器的作用。Assert Loser：从出接口列表中删除该接口。

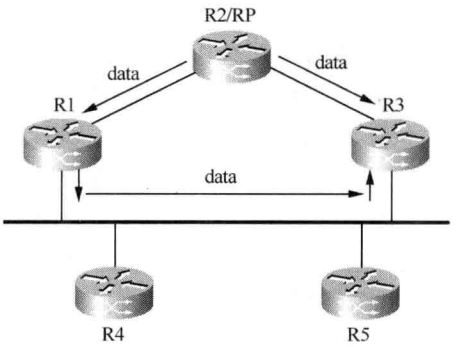


图 3-30　Assert 图示

3.4.9　BSR 和 RP 选举

一个网络内部只有一个 BSR，但可以配置多个候选 BSR（C-BSR），也可以有多个 C-RP。C-RP 通过单播消息通知 BSR。BSR 用 RP-Set 记录 C-RP 的信息，并周期性地向所有路由器发送 RP-Set 的信息。所有路由器利用同一种算法从 RP-Set 中选择出 RP。

具体来说，BSR 机制按照如下描述进行，主要有 4 个步骤，尽管实际上各个步骤是同步进行的。

① BSR 选举。每个 CBSR 组织一个 Bootstrap messages（BSMs）。每一个 BSM 都包含有一个 BSR 优先级选项，域中的路由器在该域中洪泛 BSM。当一个 CBSR 接收到一个比自己优先级更高的 CBSR 发送的 BSM 时就会抑制自己，并且在一定时间内转发高优先级的 BSM，最后那一个剩下的 CBSR 就被选举成 EBSR，并且它的 BSM 会通知域中的所有其他路由器它被选举为 EBSR。

② CRP 公告。域中的每一个 CRP 都会周期性的发送 Candidate-RP-Advertisement（C-RP-Adv）报文给 EBSR，一个 C-RP-Adv 报文包含 CRP 的优先级和 CRP 负责的组范围列表，这样 BSR 就能够获得可用的 RP。

③ RP-Set 的构造。BSR 在收到 C-RP-Adv 报文的 CRP 中选择一个子集来构成 RP-Set。通常 BSR 必须保证 RP-Set 不能太大，以便通知域中的所有路由器。RP-Set 又不能太小，否则会使负载集中在个别 RP 上，应尽量保证 RP-Set 不经常改变。

④ RP-Set 洪泛。BSR 将 RP-Set 包含在 Bootstrap messages 中。BSM 报文在域中洪泛以保证 RP-Set 能快速地到达域中所有的路由器。BSM 被周期性地发送以确保一致性。

如果一个 PIM 域被分割开来，每一个新的区域都会选举出自己的 BSR，由新的 BSR 根据可达的 RP 形成新的 RP-Set。当几个部分合并时，新的选举会自动进行，并且只会选举出

一个新的 BSR 来发送 BSM。为了使一个新的路由器尽快获得 RP-Set，当收到一个新邻居的 Hello 报文，或者收到一个已有邻居的具有一个新的 GenID 的 Hello 报文时，路有器就会单播一份自己储存的 BSM 报文的拷贝到新的或者重启的路由器。一般由网络上的 DR 负责单播 BSM 消息。

当一个 PIM 路由器接收到 BSM 消息后，会综合由其他方式得到的组和 RP 的信息，形成一个组和 RP 对应的表，并且根据特定的方式计算出组地址和 RP 地址的最终映射表，形成多播转发树。每个 PIM 域中的每个路由器都必须能够将特定多播组的地址映射到相同的 RP 上面。RP 的映射算法如下。

① 通过与多播组范围最长匹配来获得 RP 列表。

② 从这些匹配的 RP 列表中找出优先级最高的 RP，去掉其他低优先级的 RP。

③ 如果只有一个 RP 符合条件，就使用那个 RP。

④ 如果有多个 RP 符合条件，采用 PIM 的 Hash 算法选取一个 RP。

如果有多个映射表映射相同的多播组，那么采用掩码最长的那个映射表；如果掩码长度相同，则选取优先级最高的那个映射表。如果有不止一个 RP 可以映射，而且 Hash 掩码长度和优先级都相同的话，那么就采用 Hash 功能选举 RP。

3.4.10 多路访问网络的处理

在多路访问网络上运行 PIM-SM，需要一定的特殊处理机制。

① DR 选举：DR 负责为组成员向多播分发树的根节点（RP 或源）发送加入/剪枝消息，或者将直连多播源的数据发向多播分发树。

② 剪枝否决：多路访问网络上，其他路由器监听到剪枝消息时，若它仍希望接收该上游路由器发来的多播数据分组，则通过向上游路由器发送加入消息来否决剪枝。

③ 加入/剪枝抑制：多路访问网络上的路由器监听到加入消息时，启动抑制定时器来抑制加入/剪枝消息。定时器超时，才发送消息。此机制有助于减少网络上的冗余信息。

④ 断言（Assert）。

3.4.11 PIM-SSM

SSM 为指定多播源的多播技术。通过 PIM-SSM 和 MLDv2 相结合，能够在多播源和多播数据接收者之间直接建立最短路径树（SPT），省去了 PIM-SM 中先建立共享树，再从共享树向最短路径树切换的过程，从而能够从一开始就沿最短路径树转发数据。因此，与其他多播技术相比，在已知多播源的情况下，SSM 技术效率高，而且简化了多播地址分配，能够较好地应用到多多播源的情况。

PIM-SSM 通常依托 PIM-SM 实现，它可以和 PIM-SM 共存于同一台路由器。根据数据和协议报文中的多播地址来决定使用 PIM-SSM 还是 PIM-SM。IANA 为 SSM 分配了地址段 FF3x::/32。

当 SSM 接收到 MLDv2 的 IS_IN 报文，获知接收到 MLD 报文的接口所连接的网络上有主机想要接收发往多播组 G 来自源 S 的数据报后，根据单播路由向连接多播源的第一跳路由器逐跳朝源的方向发送 IPv6 PIM（S,G）源组加入报文，从而在多播源和连接接收者的最后一跳路由器之间建立起最短路径树。当多播源发送多播数据的时候，这些数据就沿着最短路

径树到达接收者。

3.5　IPv6 多播协议的现网部署策略建议

3.5.1　RP 部署

IPv6 在 RP 部署上与 IPv4 有相似之处，有静态 RP 和 BSR RP（候选者 RP）方式。除此之外，IPv6 还有嵌入式 RP（Embeded RP）方式。

1．静态 RP

这种方式是将 RP 静态配置于城域网核心设备上，可以对到 RP 的映射多播组地址做 ACL 控制，只允许指定的多播组地址通过映射。

其他域内多播路由器配置静态指定 RP 地址，通过 IGP 获悉 RP 地址的路由。

2．BSR RP

BSR RP 称为引导路由器 RP，提供穿越 PIM 域的 RP 到组映射信息的分发。选择 2 台 SR 配置成 2 台 BSR 的候选（candidate）路由器，通过优先级区分主备，优先级高的代表主 BSR，低的代表备份 BSR。选择 2 台核心路由器配置候选者（candidate）RP，优先级低的为主，高的为备。需要配置不同的 RP 地址和不同的优先级。

当部署完 candidate BSR 和 candidate RP 后，这 4 台路由器发送 BSM。候选者 RP 路由器优选出 1 台主 RP，候选者 BSR 优选出 1 台主 BSR。主 BSR 向整个域内泛洪 BSM，将 RP 信息通告给其他域内非 BSR 多播路由器，这样其他路由器就知道 RP 的地址是在哪台路由器了。

这种 RP 部署方式类似于 IPv4 中的 BSR RP 部署。

3．Embeded RP

嵌入式 RP 是 IPv6 多播提出的新概念。与 IPv4 不同，这种方式 RP 是从多播组地址中提取出来。当然多播组地址有标识位表示这个多播组地址中含有 RP 地址的前缀和 ID 号。FF7 开头的多播组地址中才含有 RP 地址，如用 2002:6666::6 作为 RP address，多播地址可以规划为 FF7y:06yy:2002:6666::yy，y 表示可以自定义。

此种方式同样需要在核心路由器上配置 loopback 口地址作为 RP 地址。其他域内多播路由器如果支持嵌入式 RP 的概念，这台路由器就会在 MLD 报告、PIM 流量或数据流量中搜索 RP 地址，搜索方式是查找多播组地址中 R 标志为 1 的目的多播地址。

3.5.2　多播路由协议的部署

一般网络中 CR 与 SR/BRAS 的互联链路两端部署 PIM-SM 协议。多播业务多播流的复制点选择 2 种测试方式：在 BRAS/SR 设备上、在接入层设备上。

3.5.3　MLD 静态加入的部署

IPv6 的多播组成员管理使用的协议与 IPv4 的类似，但协议名称不同。IPv6 采用 MLD 协议。在 BRAS/SR 部署 MLD 协议，接入网同样开启 MLD 协议。

3.5.4 二层设备开启 MLD snooping

在二层接入设备，如交换机和 XPON 设备上打开 MLD snooping，接收/转发多播流。

3.5.5 RP 备份

IPv6 多播不支持 MSDP，可考虑通过以下 3 种方式实现 RP 冗余。

① 静态 RP 和嵌入式 RP 映射方式，都需要在 2 台路由器上配置 loopback 口地址作为 RP 地址，2 台路由器上配置相同 IPv6 地址和不同 IPv6 前缀实现冗余。如，R1 上配置 2001:D:AAAA:1::2/128，R2 上配置 2001:D:AAAA:1::2/127，当 2 个 RP 都是正常运行状态时，会选择 R1；但如果 R1 宕机，则选择 R2 上配置的 RP 地址。

② BSR RP 方式，2 台路由器配置相同的 RP 地址，一主一备，根据优先级高低选择。

③ 可以通过 RFC 4610 对 PIM 的扩展方案进行部署，实现 Anycast RP。在 IPv6 多播中，没有 MSDP 协议，那么采用 PIM-SM 协议扩展的实现方式来实现 Anycast RP 功能是比较好的选择。

一个域内有多个 RP 来分担，接收者和源寻找最近的 RP。从源来的数据分组必须到达所有的 RP，寻找加入的接收者。选择一个单播的 IP 地址为 RP 地址，此地址为静态配置或者使用协议向域内所有 PIM 路由器通告。所有 RP 的对外地址都是这一个。这些 RP 被称为 Anycast-RP Set。这些 RP 在内部通信时又各自拥有一个 IP 地址。

第 4 章
IPv6 过渡技术

4.1 总体介绍

随着 IPv4 地址的逐渐耗尽，而现在很多业务却有巨大的 IP 地址需求（包括固定宽带、移动互联网、物联网等，中国电信等运营商持续发展的业务对 IP 地址有上千万甚至数亿的需求），到 2012 年左右，中国运营商将无法获得 IPv4 的地址，业务发展的需要和 IPv4 地址耗尽成为必须面临的现实问题。从长远来看，IPv6 是针对 IPv4 地址短缺问题的根本解决办法，现在的关键是如何在 IPv4 地址枯竭前完成向 IPv6 的平滑演进。如果不能在 IPv4 地址耗尽前实现向 IPv6 的平滑迁移，那么整个互联网将面临严重的 NAT 问题，运营商将迫不得已采用 IPv4 私有地址发展业务，这样将导致网络中需要部署大量的 NAT 设备，以满足 IPv4 私有地址到 IPv4 公有地址的转换，整个互联网将由此进入 NAT 时代。

IPv6 是针对 IPv4 地址耗尽问题的根本解决方案，但是因为 IPv6 与 IPv4 技术不兼容，导致从 IPv4 向 IPv6 难以实现平滑迁移。为了保证用户和业务平稳过渡，整个演进周期将是长期的过程，IPv6 会与 IPv4 长期共存。

对国内运营商来说，IPv4 公有地址储备量较少，2012 年左右将无法再从 IP 地址分配机构获得新的 IPv4 公有地址，而 IPv6 端到端整体的产业环境还无法满足采用纯 IPv6 运营时，将面临 3 个艰难的选择。

① 为新用户分配 IPv4 私有地址，在运营商网络部署传统的 NAT 设备解决 IPv4 私有地址和 IPv4 公有地址的转换，暂时不引入 IPv6。

② 为新用户分配 IPv6 单栈地址发展业务。为了使这些用户访问 IPv4 业务并与现存的 IPv4 用户之间进行通信，必须部署 IPv6 与 IPv4 互通转换设备。

③ 为新用户同时分配双栈地址（IPv4 私有地址 + IPv6 地址），访问 IPv6 业务用 IPv6 地址，访问 IPv4 业务用 IPv4 私有地址。运营商网络部署传统的 NAT 设备以解决 IPv4 私有地址和 IPv4 公有地址的转换问题，IPv4 与 IPv6 之间不进行互通。

哪种方式更适合运营商？哪些过渡技术可以在其中发挥重要作用？就这些问题，华为和国内运营商一直在紧密合作，不断进行探索，逐渐找到适合国内运营商的 IPv6 过渡技术方案。

通常所说的 IPv6 过渡技术，一般分为 3 大类：双栈、隧道、地址转换或翻译技术（NAT）。

支持 IPv4 公网、IPv4 私网和 IPv6 共存，以及平滑过渡是网络演进的关键技术。支持

IPv4 私网和公网共存的技术是 NAT。支持 IPv4 向 IPv6 过渡的技术（过渡技术必然支持 IPv4 和 IPv6 共存）有很多，发展至今被公认为适合运营商部署的只有双栈、DS-Lite 以及演进后期的 NAT64（也是 NAT 技术）。6RD 技术曾被寄予厚望，但试验后发现不适合运营商环境，主要原因如下。

① 缺乏产业链支持。

② 完全依赖现有 IPv4 网络，后续向 IPv6 演进需要二次网改。

③ IPv6 通过隧道封装在 IPv4 网络上，通过独立的 6RD 网关解封装，无法管理到用户和业务，QoS 和流量管理难以实施，导致网络不可管理。

④ 桥接式 CPE 需 PC/终端安装特定驱动程序，导致低成本的桥接式 CPE 很难广泛实施；而路由型 CPE 需要支持 6RD，与其他方案相比又无成本优势。

双栈方案和 DS-Lite 方案都是基于双栈的，区别在于 DS-Lite 允许运营商只部署 IPv6 单栈网络，但是参与通信的终端必须同时为 IPv4 或者同时为 IPv6。只有 NAT64 才支持 IPv6 终端访问 IPv4 网络服务。

4.2 双栈技术

双栈的定义在 RFC 4213 中，是指在终端设备和网络节点上既安装 IPv4 又安装 IPv6 的协议栈，从而实现分别与 IPv4 或 IPv6 节点间的信息互通。具有 IPv4/IPv6 双协议栈（如图 4-1 所示）的节点简称双栈节点，这些节点既可以收发 IPv4 报文，也可以收发 IPv6 报文。它们可以使用 IPv4 与 IPv4 节点互通，也可以直接使用 IPv6 与 IPv6 节点互通。

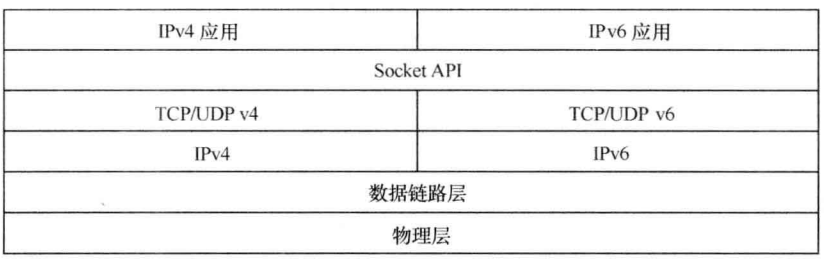


图 4-1 IPv4/IPv6 双协议栈结构图

配置为双栈的设备接口可以拥有一个 IPv4 地址或一个 IPv6 地址或两者都拥有。同时，路由器包含两个单独的路由表：一个用于 IPv4 寻址，另一个用于 IPv6 寻址。两个表都驻留于同一路由器中。

当双栈节点链路层接收到数据段时，拆开并检查分组头。如果 IPv4/IPv6 头中的第一个字段，即 IP 分组的版本号是 4，该分组就由 IPv4 栈来处理；如果版本号是 6，则由 IPv6 栈处理。

为了支持 IPv6 路由的学习，双栈路由器设备还要同时运行支持 IPv6 的路由协议。如果现网络部署了 OSPF，则新增支持 IPv6 的 OSPFv3；如果现网络部署的是 ISIS，则在网络中部署 ISIS 多拓扑，以支持 IPv6 路由的学习。支持 IPv6 的 BGP4 +，通过配置使能 IPv6 地址族来支持 IPv6 路由的通告，必要时升级 RR 以支持 IPv6 路由反射功能。

双栈体系结构允许设备接收、处理和转发 IPv4/IPv6 信息流。支持 IPv4/IPv6 双栈的网络

设备（路由器等），使网络从逻辑上看到并行的两个网络，并支持向 IPv6 的平滑过渡。

双栈机制是使 IPv6 节点与 IPv4 节点兼容的最直接的方式，其互通性好，易于理解。但是双栈并不能解决 IPv4 地址紧缺问题，所以还需要引入 NAT 技术。

4.2.1　IP 骨干网

IP 骨干网支持双栈，主要有两种技术选择：Native 双栈和 MPLS 6PE/6vPE，两种技术比较见表 4-1。

① Native 双栈是在现有的 IPv4 路由器和链路上，使能 IPv6 和路由协议。IPv6 报文与 IPv4 报文一样，直接封装在链路层上转发。

② MPLS 6PE/6vPE 则是保持现有的 MPLS 网络不变，在 PE 上将 IPv6 报文封装到 MPLS 中进行转发。两种方式都有广泛的实际部署应用。

表 4-1　　Native 双栈与 MPLS 6PE/6vPE 比较

	Native 双栈	MPLS 6PE/6vPE
实施成本	所有设备使能 IPv6，实施成本稍高	PE 使能 IPv6，P 不必使能 IPv6，实施成本较低
协议部署	所有路由器运行 IPv6 IGP/BGP	MPLS 核心不变，PE 之间部署 MP-BGP
可扩展性	没有限制	没有限制
维护成本	需在所有节点维护新引入的IPv6协议和路由	IPv6 作为 MPLS 新业务，对现有维护冲击较小，维护范围限制在 PE
业务支持	单播/多播	多播不成熟，可支持 VPN 应用

有观点认为，双栈对设备资源和维护的要求高，不是合适的解决方案。但对骨干路由器而言，本身就是高性能设备，且路由器专用的转发和查找芯片技术仍然在快速发展，未来并不存在资源瓶颈。至于维护，网络发展的历史已经充分说明，维护能力和资源应当匹配新业务新技术的发展。所以骨干网的演进路线很可能就是双栈，直到最后 IPv4 网络关闭。

4.2.2　BNG

IP 边缘节点统称为 BNG（Broadband Network Gateway），BRAS 和 SR 都属于 BNG 的表现形式。BNG 主要用于用户和业务控制，包括分配地址。

但 BNG 支持双栈不仅仅是为用户同时分配 IPv4 地址和 IPv6 地址，在会话管理、计费、QoS、安全等方面都需要修改以适应 IPv6 业务。由于 BNG 往往与网管、AAA、DHCP、Policy 等业务系统有接口，因此 BNG 支持双栈还需要这些周边系统配合，在导入 IPv6 之前应该进行详细的集成测试。

IPv6 地址分配方式与 IPv4 有较大的差异，由于 IPv6 地址空间巨大，推荐为用户分配一个合适的网段前缀（参考 RFC 3633），这也是路由型 CPE 将占主流的原因。出于管理需要，建议为 CPE WAN 接口分配一个单独的/128 主机地址。

4.2.3　CPE

CPE 终端是网络演进部署的关键。无论采用 PPPoE 还是 IPoE 架构，长期来看，路由型双栈 CPE 将是主流。

CPE 采用 DHCPv6 Prefix Delegation 从 WAN 接口向 BNG 获取 IPv6 前缀，并将其子网化，分配给 LAN 接口。在 LAN 侧，用户终端可以通过 DHCPv6 或 SLAAC 方式获得地址，CPE 应该同时支持这两种模式，因为不同用户终端的能力可能存在相当大的差异。CPE 应具备升级到支持 IPv6 多播的能力。目前对于家庭网络的 IPv6 多播架构还不是很完善，CPE 应保持硬件 Ready 状态，使得通过软件升级即可支持 IPv6 多播业务。未来高清电视、互联网视频对带宽的要求很高，CPE 应该对任意比例的 IPv4、IPv6 混合业务具备高速转发能力。

CPE 需要完全保留原来 IPv4 的功能，如 DHCP、NAT、端口映射、UPnP 等。随着 NAT 技术的发展，CPE 应具备可升级能力，支持新协议部署。

CPE 应当支持通过远程软件升级或修改配置即可支持不同的 IPv6 业务部署方案。向 IPv6 网络演进中存在调整部署方案的需求，比如早期部署 6RD 或双栈，后期由于 IPv4 地址不足而切换到 DS-Lite，需要在无入户的前提下就可支持演进技术方案的迁移，以降低部署成本。

4.3 隧道技术

隧道技术是将一种协议的数据报文封装在另一种协议的数据报文中传输的技术，它是连接孤岛的有效方法。隧道类型有很多种，包括 6PE、6vPE、手工隧道、GRE、L2TP、6over4、4over6、6to4、ISATAP、Teredo 等，不同类型的隧道可以用于不同的应用场景。隧道技术有一个较大的问题：需要两端设备之间能够良好互通。

4.3.1 6PE

6PE（IPv6 Provider Edge）是在目前的 IPv4 网络中利用 MPLS 隧道技术为不同地区的被分割的 IPv6 网络提供连通服务。将被分割的 IPv6 网络利用隧道技术连接起来的方式有很多。6PE 方式的隧道是在 ISP 的 PE 设备上实现 IPv4/IPv6 双协议栈，利用 MP-BGP 为其分配的标签标识 IPv6 路由，并通过 PE 之间的 LSP 实现 IPv6 之间的互通。

6PE 要求运行在 ISP 网络边缘，与用户 IPv6 网络连接的 PE 设备采用 IPv4/IPv6 双协议栈，PE 和 CE 之间利用 IPv6 协议中的 IGP、EBGP 和静态路由等方式交换 IPv6 路由。PE 和 P，以及其他 PE 之间利用 IPv4 路由协议交换路由。PE 之间需要创建 Tunnel 来透明传输 IPv6 报文。

使用 6PE 作为 IPv6 网络间连通的隧道的主要优势在于以下几方面。

① 所有配置在 PE 上完成，用户网络感知不到 IPv4 网络的存在。

② 能够很好地利用 ISP 现有的 MPLS 网络资源，对运营商网络改造不大。

③ PE-CE 之间的链路可以使用任何类型，没有特殊要求。

④ 6PE 设备可以同时为用户提供 IPv6 VPN 和 IPv4 VPN 等多种业务。

在 6PE 组网中，默认情况下，6PE 路由是每路由每标签方式，即每条发送到其他 6PE 对等体的路由都需要申请一个标签，占用标签的数量与需要发送的 6PE 路由数成正比，当要发送大量 6PE 路由时就会占用大量的标签资源。

使能 6PE 路由共享显示空标签后，所有发往其他对等体的 6PE 路由会共享显示空标签 2，而不用再为每条路由申请标签，占用标签的数量与 6PE 的路由量无关，这样就大大节省了 6PE ME 设备上的标签资源。

显示空标签 2 是一种特殊标签，表示该标签在 Egress PE 上必须被弹出，且报文的转发

必须基于 IPv6。

需要注意的是，在 6PE 对等体已经建立的情况下，使能和去使能 6PE 路由共享显式空标签的时候会存在短暂的分组丢失现象。所以建议在一开始建立 6PE 对等体时，就使能该特性。

4.3.2　6vPE

在 BGP/MPLS IP VPN 中，PE 和 PE 之间、PE 和 CE 之间都运行 IPv4 的路由协议，如 BGP、OSPF、ISIS 等。当 VPN 客户网络从 IPv4 演进到 IPv6 后，PE 与 CE 之间运行上述路由协议将不再适用，在骨干网上传送的将是 IPv6 VPN 报文。BGP/MPLS IPv6 VPN 扩展使得 VPN 骨干网不必升级到 IPv6 网络，就可以给客户提供 IPv6 的 VPN 服务。

在提供 IPv6 VPN 服务的同时，PE 和 PE 之间的运营商骨干网仍然运行 IPv4 协议。这样，可以使运营商网络逐步从 IPv4 过渡到 IPv6。

骨干网是 IPv4 的情况下，PE 之间使用 IPv4 地址建立 VPNv6 邻居通告 VPN-IPv6 路由，VPN-IPv6 路由可以选择骨干网中的 IPv4 隧道来承载 IPv6 VPN 业务。

BGP/MLS IPv6 VPN 除 PE 和 CE 之间运行的路由协议与 IPv4 VPN 不同外，其他所有特性原理都与 IPv4 相同。

4.3.3　6over4 隧道

6over4 隧道方案是基于 IPinIP 技术，将 IPv6 报文封装到 IPv4 报文中，在现有的 IPv4 网络中传输的一种过渡技术。

与 6PE/6vPE 类似，6over4 隧道方案不需要骨干网络升级来支持 IPv6，而与 6PE/6vPE 不同，6over4 隧道方案不需要 MPLS 支持。隧道方案部署相当简单：按需开启/升级路由器来支持 6over4 功能即可。

图 4-2 所示为局部部署的案例，远端的路由器通过 IPv4 网络接入 IPv6 路由器。

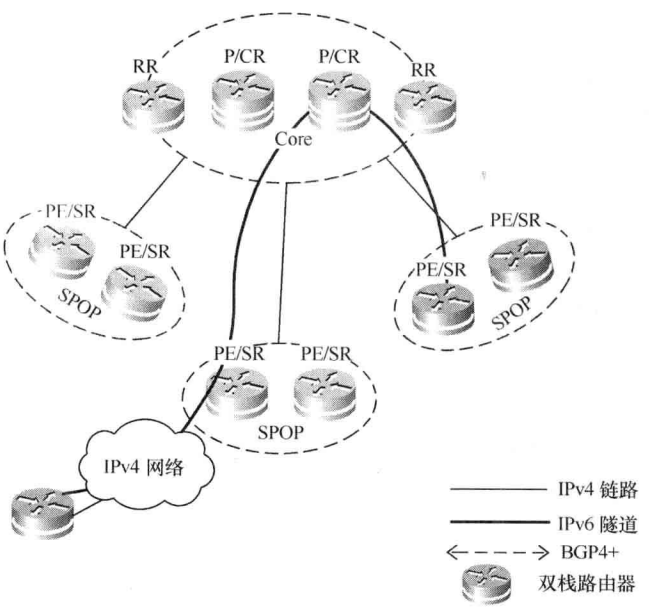


图 4-2　6over4 隧道部署示意图

6over4 报文转发模式如图 4-3 所示。

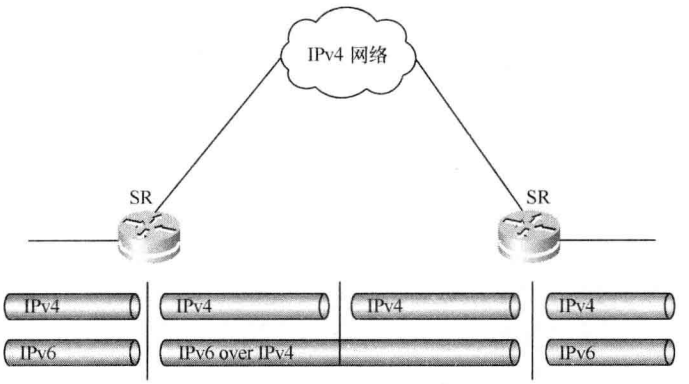


图 4-3　6over4 转发模式示意图

6over4 隧道技术在引进 IPv6 的早期阶段也可用于全网部署，但网络运营商应当了解这种部署的局限。全网部署隧道技术时，在 IPv6 路由设计上应当尽量设计为单层结构，所有 IPv6 路由器通过 6over4 虚拟链路一跳直连，避免隧道链路之间做路由转发。隧道技术下做网络分层结构是没有意义的。

1．6over4 隧道适用场景

手工隧道适用于临时过渡只需要提供可达性的场景，比如计划开通 IPv6 的区域暂时无法获得 Native IPv6 连接，则可通过隧道连接到现有 IPv6 路由器。

大规模在运营商环境部署 6over4 隧道，或者采用隧道作为过渡方案是不合适的，原因如下。

① 需要手工隧道的数量为 $2N$，配置和维护都比较困难。

② 传输效率较低，标准规定只允许隧道 MTU 为 1280。

③ 不能提供 QoS/多播业务。

④ 隧道面临安全问题，比如放大攻击、难以被防火墙防护等。

特定场合下，通过 VLL/VPLS 为 IPv6 提供虚拟二层连接以开通 IPv6 也可能是可行的简单替代方案。

2．隧道类型

用于路由器间穿越 IPv4 网络的 6over4 隧道方案有如下类型。

① IPv6 over IPv4 手工隧道。

② IPv6 over IPv4 GRE 隧道。

优选 IPv6 over IPv4 手工隧道方式，因为这种方式不需要隧道 SPU 板，而 GRE 需要 SPU 板。IPv6 over IPv4 手工隧道为硬件转发。

3．6over4 隧道的路由部署

隧道部署下，隧道两端节点是双栈路由器，中间节点只需要支持 IPv4 即可。这样，隧道路由器便将 IPv4 网络作为点到点链路来使用。

在点到点链路上可以允许 IPv6 路由协议，比如 OSPFv3 或 BGP4 + 。IS-IS 由于是链路层的协议，不能在隧道上运行。

由于隧道部署是 overlay 网络，如果要部署 IPv6 动态路由，则应该尽量简单部署，比如使用单跳连接方式，所有 6over4 路由器都利用 IPv4 网络直连。过于复杂的分层部署可能导致流量绕行，难以预测和管理。

图 4-4 所示为隧道部署案例，6over4 路由器包括 IGW 和 PE，它们通过隧道直连。在隧道上运行 OSPFv3 协议发布环回地址，BGP4 + 基于环回地址建立邻居关系进行业务路由的学习。

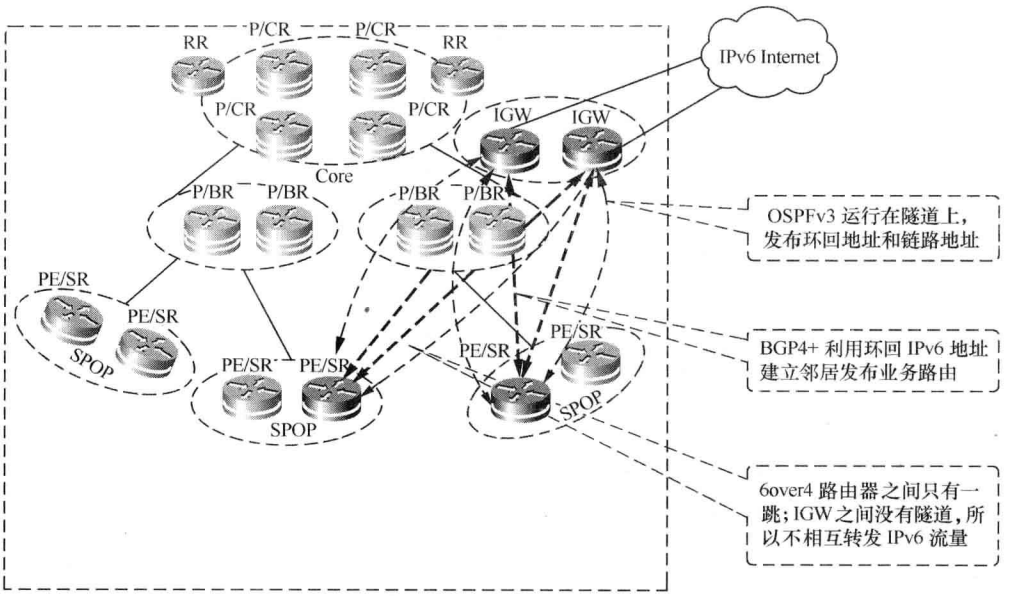


图 4-4　三级汇聚结构下部署 6over4 隧道案例图

4.4　协议翻译技术

协议翻译技术统称为 NAT（Network Address Translation）技术，可以分为两类，一类是 IPv4 私有地址到 IPv4 公有地址的地址转换 NAT 技术(或称为 NAT44)，另一类是 IPv4 与 IPv6 地址之间转换的技术，例如 NAT-PT、NAT64、IVI 等。

4.4.1　NAT44 技术

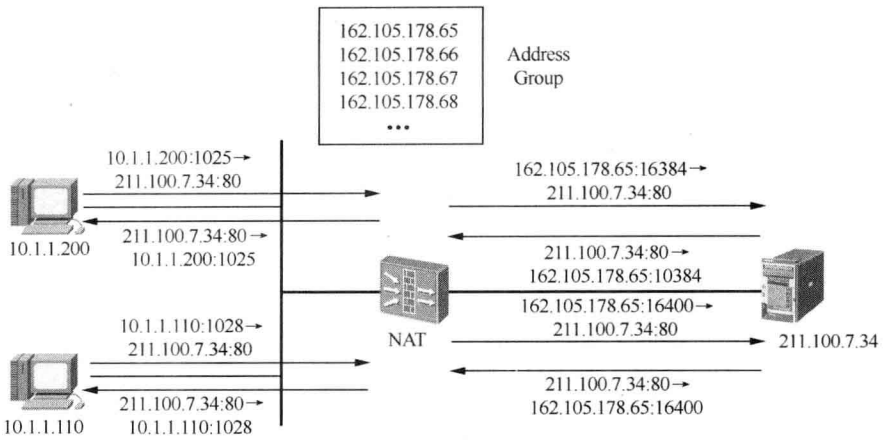
1. NAT 原理

地址转换有两种方式：NAT 和 NAPT（Network Address and Port Translation）。NAT 方式只转换 IP 地址，对端口号不处理。而采取 NAPT 方式时，NAT 设备进行地址转换的时候，不仅要对 IP 报文中的地址进行转换，还要对报文中的端口号进行转换。但应用 NAT 时，每个私网 IP 都需要一个公网地址，对地址比较浪费。因此，实际应用中多采用 NAPT 方式。

图 4-5 所示为 NAPT 方式的基本原理，其对报文的处理过程如下。

① NAT 设备接收到私网用户发出的访问公网的报文。

② 如果是私网用户对外发起一个新连接，NAT 设备会从地址池中选择一个空闲的公网地址和端口号、建立 NAT 转换表项。



NAT 表	NAT 前	NAT 后
outbound	10.1.1.200:1025	162.105.178.65:16384
inbound	162.105.178.65:16384	10.1.1.200:1025
outbound	10.1.1.110:1028	162.105.178.65:16400
inbound	162.105.178.65:16400	10.1.1.110:1028

图 4-5　NAPT 方式基本原理

③ 根据私网源 IP 地址、目标 IP 地址、源端口号和目标端口号查找 NAPT 表项。根据查表结果转换报文，并向公网侧发送。

④ NAT 设备接收到公网侧的回应报文，根据目的 IP 地址和端口号查找反向 NAPT 表项。根据查表结果转换报文，并向私网侧发送。

NAPT 方式同时转换 IP 地址和端口号，可以更加充分地利用 IP 地址资源，实现更多的内部网主机对 Internet 的同时访问。

2. NAT 报文处理流程

从私有网络到公有网络的报文在处理过程中，接口卡会针对所有的流量都开启 ACL。现在的 ACL 一般都是使用 TCAM 查找，不影响转发性能，可以线速转发，NAT 报文处理流程如图 4-6 所示。

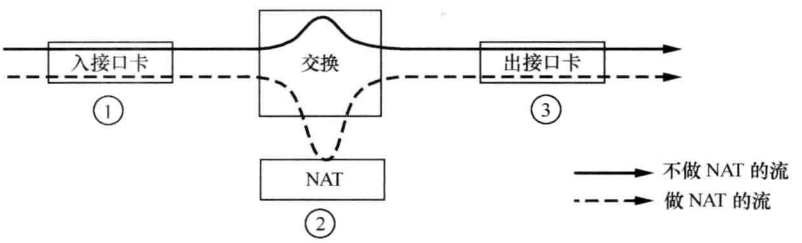


图 4-6　NAT 报文处理流程

① 在入接口卡查找 ACL/UCL 表，发现是需要 NAT 转换的流量，将流量发送到 NAT 卡。

② 在 NAT 卡上根据三元组（源 IP + 源端口 + 协议号）或五元组（源/目的 IP + 源/目的端口 + 协议号）查 IPv4 正向会话表，NAT 卡对报文进行 NAT 转换（转换报文的 IP 地址和端

口号等）；查 FIB 表转发到接口卡下行继续转发到网络侧出口。五元组匹配条件较三元组严格，在考虑安全等因素时使用五元组。

③ 出接口卡，对已经完成 NAT 转换的流量转发流量。

对于回程流量，其报文目的 IP 地址是 NAT 上公网 IPv4 地址，报文被路由到 NAT 单元并做反向 NAT 处理。

3．NAT 技术需求

（1）性能容量

运营商环境下，NAT 支持的用户数非常多，可能达到十万级别的用户数，每个用户的平均流量可能达到几百 kbit/s，NAT 设备需要 100Gbit/s 级别的转发能力。实验表明，Web 2.0 网页的点击会生成数十个 TCP 连接，P2P 应用会生成超过 100 个会话，每个用户预留 1000 个端口配额一般可够用，NAT 设备需要具备每秒新建百万会话、维持千万活动会话的能力。

（2）可靠性

可通过部署 NAT 设备冗余和设备内板级备份来提高 NAT 网络的可靠性，这样当设备出现故障后能够自动切换到备用设备上。与一般业务不同，NAT 会话是有状态的。会话的生成和老化非常快，NAT 上可能会达到每秒百万会话的状态变化，而备份会话需要进行协议交互，在这种情况下备份几乎不可能做到可靠，而且一般认为绝大部分 NAT 会话的生存时间都极短，备份价值非常小。所以目前被广泛接受的是只对生存时间较长的会话进行备份。

（3）用户管理

NAT 应用环境下需要对用户进行端口配额管理，避免少数用户滥用网络而导致其他共享 IP 的用户无法正常使用网络。此外，一般还需要保持外部地址的唯一性和端口的奇偶性，规避某些特殊端口（例如可能被判别为病毒的端口）等，需要 NAT 具备可管理的特性。

（4）地址溯源

部署 NAT 还必须考虑溯源问题，而溯源需要应用层面的支持，比如网站的访问日志不能仅仅记录 IP 地址，还需要记录端口信息。按照普通实现 Session 记录日志的方法，在 NAT 环境下，日志流量可能高达数十 MB/s，需要极高性能的日志处理和存储系统，提高了运维成本。NAT 可通过支持端口预分配技术，一次为用户预留数百上千个端口，从而降低 NAT 日志规模至千分之一或更小。

（5）NAT 穿越

TCP/UDP 的有些应用（例如多媒体会话、文件共享、游戏等“端到端”的应用）中，报文的负荷部分有 IP 地址或者端口的信息。NAT 做地址转换时，如果不处理负荷中的 IP 地址或者端口信息，应用程序会失败。

4．NAT 应用场景

NAT 与其他技术结合有多种应用场景，如 NAT444、DS-Lite、NAT64 等。

① NAT444 包含两次 NAT，分别在 CPE 和运用商网络中的 NAT 网关上做地址转换。

② DS-Lite 场景引入了两个新概念：AFTR（Address Family Transition Router）和 B4（Basic Bridging BroadBand element）。AFTR 是 DS-Lite 网关，负责隧道的封装与解封装，以及地址转换；B4 相当于 CPE 的位置，承担隧道的封装与解封装。CPE 和 AFTR 之间配置成单栈 IPv6 网络，B4 向用户分配私有 IPv4 地址，以支撑 IPv4 业务。IPv4 报

文上行时由 B4 将 IPv4 报文封装在 IPv6 隧道中。到达 DS-Lite 网关时，解除外层 IPv6 封装并完成私有 IPv4 地址和公用 IPv4 地址的转换。来自公网的 IPv4 报文的处理过程则相反，先将公用 IPv4 地址转换成私有 IPv4 地址，并由 IPv6 隧道送达 B4 设备，B4 进行解封装并转发给目的主机。

③ NAT64 同时实现 IPv6 与 IPv4 之间的网络地址与协议转换技术。NAT64 一般只支持 IPv6 网络侧用户发起连接访问 IPv4 侧网络资源；如通过手工配置静态映射关系，NAT64 也支持 IPv4 网络主动发起连接访问 IPv6 网络。NAT64 可实现 TCP、UDP、ICMP 协议下的 IPv6 与 IPv4 网络地址和协议转换。DNS64 要配合 NAT64 工作，将 DNS 查询信息中的 A 记录（IPv4 地址）合成到 AAAA 记录（IPv6 地址）中，并将返回合成的 AAAA 记录给 IPv6 侧用户。

5. NAT 对网络的影响

（1）对网络设备的影响

部署 NAT 需要在网络中增加地址转换功能，对现网调整不大，部分应用系统需要针对私网地址作修改。NAT 不仅仅增加了处理时延，还增加了网络和路由的复杂性，而且 NAT 本身是流量的汇聚点，每 session 的备份在运营商环境中难以实施，一旦发生故障可能需要终端用户干预以重新建立 session，从而降低了网络的可靠性。

（2）对网络维护的影响

用户溯源需要配置 NAT 日志服务器，以记录用户网络访问情况，这样可能导致每次 NAT 产生数十 MB/s 的日志流量，需要高性能大存储的日志服务器。增加 NAT 设备后，故障定位难度增加、用户申告以及溯源的遵从导致网络维护难度和工作量增加。

（3）对业务和应用的影响

NAT 引入的处理时延，以某些应用 NAT 穿越困难等，都会降低用户业务体验，甚至会影响运营商业务开展。例如，如果 NAT 部署于用户与 DPI 之间，DPI 功能将失效。某些运营商增值业务需要基于 IP 地址获取用户信息，NAT 也会对这类应用产生影响，因为现在地址为多用户共享；对 IP 语音通信类的应用也往往需要增加应用层网关。

尽管有这些不利影响，NAT 仍然被广泛商用部署，主要原因是地址短缺已经大范围存在，而 NAT 是唯一可保护现有 IPv4 投资的技术。支持 NAT 已成为新开发应用的必选功能，NAT 相关标准在不断发展完善，新开发的应用可以利用这些技术，通过增加应用层的复杂度来规避网络层的限制。

6. NAT 形态分类

按 NAT 设备的存在形式可以分为独立式和集成式。

① 独立式 NAT 是一个独立的设备，专门承担 NAT 功能。

② 集成式 NAT 也称插卡式，即承担 NAT 功能的一个单板插在其他功能的设备上，如插在 BNG、SR、CR 上。

以 BNG 旁挂独立 NAT 设备和 BNG 集成式 NAT 作对比，独立式 NAT 专门完成地址翻译工作，对 BNG 业务无影响，但是需占用 BNG 转发端口（引流至 NAT）；集成式 NAT 与 BNG 的用户管理相结合，可以提供更好的用户 NAT 管理能力，详见表 4-2。

表 4-2　　独立式 NAT 与集成式 NAT 的对比

	独立式 NAT	集成式 NAT
特点	性能强，多核 CPU 处理，整机全部用于 NAT 处理。 容量和扩展性较高，可通过级联或多框进行扩展。 设备安放较自由。 需占用 BNG 的转发端口向独立式 NAT 引流。	无需单独设备，空间占地小，功耗散热要求低，投资小。 网络层与应用层配合能力强，支持各种路由协议，融合接入和 NAT 管理信息。 扩展性好，可在线部署 NAT，无扩展限制。 需要占用 BNG 的业务槽位。

按 NAT 设备的部署位置可以分为集中式和分布式。

① 集中式 NAT 的部署位置通常是在网络中比较高的 CR 位置，如国内的城域网 P 路由器、海外骨干网的 P 路由器。可以是独立式 NAT 在 CR 旁，也可以是 NAT 卡集成到 CR 上。

② 相对于集中式的网络位置高而言，分布式 NAT 的网络位置较低，通常部署在 BNG 或 SR 所在的位置，可以是独立式 NAT 或集成式 NAT。

集中式 NAT 和分布式 NAT 的对比见表 4-3。

表 4-3　　集中式 NAT 与分布式 NAT 的对比

	集中式 NAT	分布式 NAT
特点	适用于整个城域网 NAT 用户比较分散的场景。 易于部署，增加节点少。 对设备性能要求高，设备故障影响范围大。 城域网内，SR/BNG 以上需要引入私网路由。 可持续演进性差，随着 NAT 用户量加大，将来还要演进到分布部署。 集中式可以减少地址池碎片，提高地址的利用率。	特别适合于 NAT 用户集中在某个区域。 NAT 用户分散情况，部署点较多。 设备故障影响范围小。 同 SR/BNG 融合后，SR/BNG 同时具有用户信息和 NAT 信息，可以基于用户来管理 NAT，方便解决溯源等由 NAT 引入的问题。

7．NAT 接入场景分类

（1）方式一：二层隧道接入，NAT444。

如图 4-7 所示，CPE 作为 IPv4 私网地址 DHCP Server，为终端网络用户分配 IPv4 私有地址。CPE 本身做一级 NAT，NAT 网关做第二级 NAT 处理。CPE 和 NAT 之间通过二层隧道（VLAN/PPPoE）连接。二层通道内仅有一个 IP 会话，二级 NAT 中的私网地址可以采用共享地址。

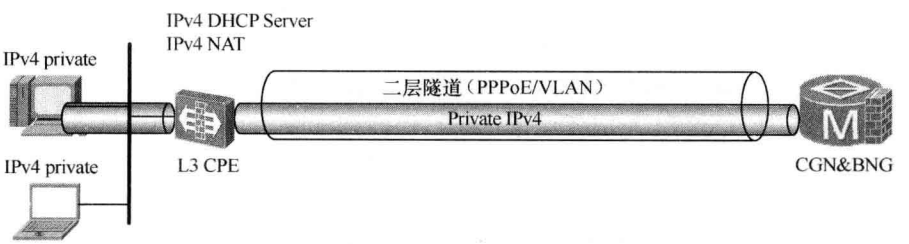


图 4-7　NAT 二层隧道接入，NAT444

（2）方式二：二层隧道接入，一级 NAT，任意地址接入。

如图 4-8 所示，终端用户通过二层隧道（VLAN/PPPoE）接入。用户 IPv4 报文通过二层

通道方式传递到 NAT 网关做 NAT 处理。CPE 作为 IPv4 私网地址 DHCP Server，为用户终端分配 IPv4 私有地址。CPE 本身不做 NAT，所有的终端网络的私网地址报文直接通过二层方式接入。二层通道内具有多个 IP 会话，终端网络用户通过二层通道信息来区分标识，如 PPP Session ID/VLAN 等。

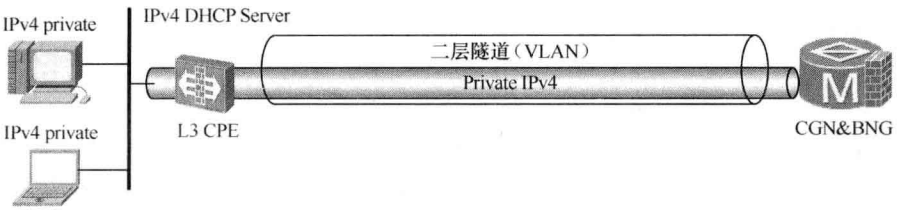


图 4-8　二层隧道接入，一级 NAT，任意地址接入

（3）方式三：二层隧道接入，一级 NAT，地址管理分配。

如图 4-9 所示，终端用户通过二层隧道（VLAN/PPPoE）接入，用户的 IPv4 报文通过二层通道方式传递到 NAT 网关做 NAT 处理，CPE 不具备地址分配能力，BNG 作为 IPv4 私网地址 DHCP Server，给用户终端分配 IPv4 私有地址。CPE 不做 NAT，所有的终端网络的私网地址报文直接通过二层方式接入。二层通道内具有多个 IP 会话，终端网络用户通过二层通道信息来区分标识，如 PPP Session ID/VLAN 等。

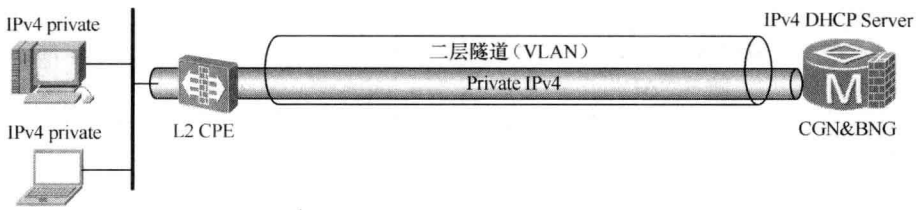


图 4-9　二层隧道接入，一级 NAT，地址管理分配

8．BNG + NAT 融合的 NAT 用户策略控制

电信级 NAT 需要做到 NAT 资源的可管可控，以及基于用户的精细化策略控制，实现地址和 NAT 资源的电信化运营分配。NAT 的管理体系架构一般以基于源地址网段/NAT 实例等做到针对用户组的策略控制，而 BNG 的管理体系架构一般以用户账号（单一用户）或控制域（用户组）进行管理。将两者融合，实现 BNG + NAT 的控制体系架构一致，可在目前 BNG 接入的网络体系架构下，以最好的兼容性实现用户的 NAT 控制策略引入。

如图 4-10 所示，NAT 控制策略包括 NAT 端口分配策略/端口段范围、NAT 会话数、合法监听策略、ALG 能力、QoS/SLA 服务质量、ACL 访问列表等。NAT 控制策略管理通过 Profile 控制策略实例来管理。控制策略 Profile 内容可以在设备本地进行配置，也可以通过策略服务器或者认证系统 Radius 下发。

针对单一用户，NAT 可以通过唯一性的用户标识来区分用户，并且在用户通过 BNG 接入的认证鉴权的时候，和用户标识绑定。NAT 控制策略可以通过用户账户策略作用于接入用户，每个用户维护用户管理控制表项，实现用户接入标识和控制策略的绑定关系。BNG/NAT 用户接入时，通过认证流程进行权限认证，获得用户的 NAT 策略，可以通过 Radius 或者策

略服务器下发给 BNG/NAT 设备，控制策略下发可以将策略的内容直接下发给设备，也可以下发控制策略模板名称，然后在设备本地配置的 Profile 实例中获得策略内容。

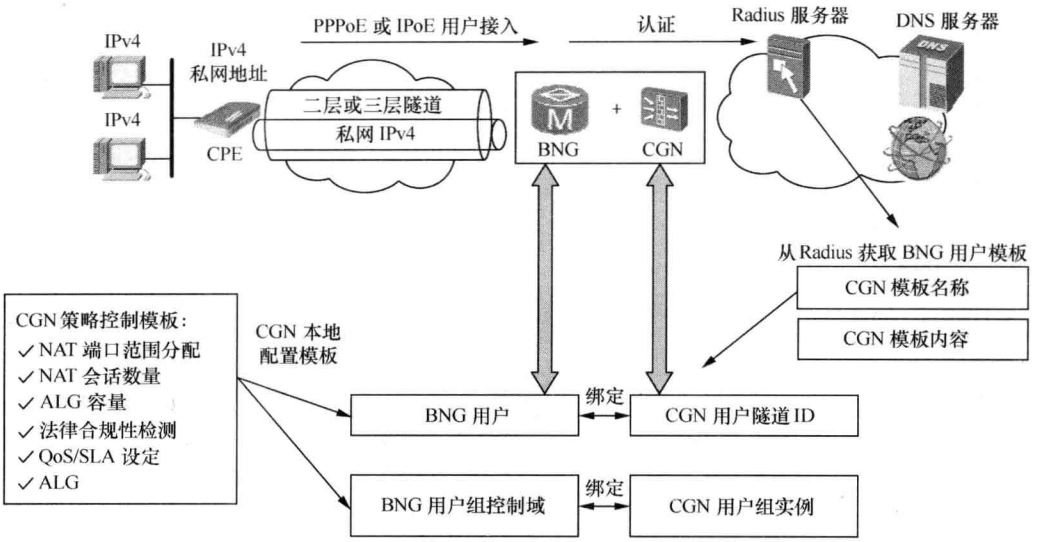


图 4-10　BNG 与 NAT 融合的用户策略控制

针对用户组，NAT 通过将配置的 NAT 实例和 BNG 的控制域进行关联。BNG 特定用户属性的用户归属于统一的控制域。控制域绑定相应的 NAT 实例，实现相应的用户组和控制策略的绑定关系。BNG/NAT 用户接入时，通过认证流程进行权限认证，通过控制域的归属关系找到绑定的 NAT 实例，获得相关联的控制策略。

NAT 策略架构融入 BNG 用户策略架构，能提供可运营的 NAT 管理策略控制能力，实现用户的地址和端口资源的有序化管理和分配，最大化地实现资源的有效利用，并且用户差异化的 NAT 业务运营和 NAT 策略统一控制集中管理，可有效地降低运维成本，实现电信级的 NAT 运营能力。

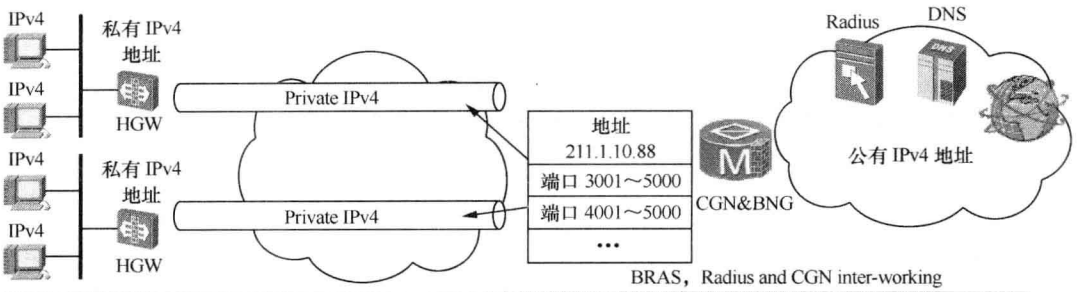
9. 有序化端口，预分配管理

通过多个用户共享 IP 地址可以实现 IP 地址的统计复用，最大化地利用有限的公网地址资源。传统的 NAT 机制一般为按需分配，先来先得。这样的分配机制造成多个管理上的问题。如，用户的端口资源占用缺乏合理控制，少部分用户可能会占用大量公网端口；端口分配耗尽后需改变 IP 地址，造成业务访问问题；用户的地址端口分配不连续，会造成用户溯源困难，需要大量的日志来记录用户的端口分配情况，造成系统负荷增加。

通过有序的端口预分配机制，可以有效地提高地址端口的使用效率，提高管理运营效率。在用户接入网络时，根据用户绑定的控制策略分配私网地址，同时指定分配给用户的公网地址和端口范围，并将所有信息记录在用户管理控制表中。

如图 4-11 所示，两个 HGW 分别分配私有 IPv4 地址——192.168.1.1 和 192.168.1.2。两个 HGW 共享同一个公有 IPv4 地址——211.1.10.88。通过不同的 TCP/UDP 端口段区分是哪个 HGW 的报文。公有 IP 地址 211.1.10.88 加端口 3001～4000 和 6001～7000 的报文对应私有

地址 192.168.1.1，公有 IP 地址 211.1.10.88 加端口 4001～5000 和 7001～8000 的报文对应私有地址 192.168.1.2。



User Identify	Private IPv4	Profile	Public IPv4	起始端口 1	终止端口 1	起始端口 2	终止端口 2	...
Tunnel ID1	192.168.1.1	Profile 1	211.1.10.88	3001	4000	6001	7000	...
Tunnel ID2	192.168.1.2	Profile 2	211.1.10.88	4001	5000	7001	8000	...

图 4-11　有序化 NAT 端口，预分配管理

用户通过私网地址访问业务时，NAT 会根据用户已分配到的公网地址和端口实现地址的翻译转换。用户的识别通过用户接入管道来唯一标识。属于同一用户的所有终端需共享该用户分配的地址和端口资源，受限于该用户的 NAT 控制策略。

给用户分配公网 IPv4 地址的外部端口，按端口范围规则有序地分配给用户，也可分配给用户组。如果端口范围内端口使用完，可根据实际需要追加分配端口给终端用户。

端口范围分配机制实现了将 IP 地址和端口资源的有序化分配管理，限制少量用户过度消耗地址和端口资源，易于对用户实现溯源追踪管理，并且大幅减少了 NAT 海量会话的日志存储量，降低了系统负荷压力。

10. NAT 穿越

NAT 有效地解决了 IP 地址资源的问题，但同时也带来了地址转换问题。NAT 可以实现 IP 层地址及 UDP 或 TCP 头端口的转换，但某些应用在 TCP/UDP 负载中也带地址信息，通常的标准 NAT 设备并不修改 TCP/UDP 负载中的内容。这样经过 NAT 后，报文的 IP 层地址与 UDP/TCP 负载中地址信息不一致，应用程序就会出错，FTP、RTSP、SIP、DNS 等协议都存在这个问题。

针对这一问题有两种解决办法。一种是 ALG（Application Layer Gateway）。ALG 是 NAT 设备具备特定应用协议的解析能力。应用程序在负载中填写的自身地址，可以被 NAT 的 ALG 模块修改为 NAT 外部地址，这样就保证了报文的 IP 报头与应用负载中地址的一致性，应用程序就不会出错。另一种方法的思路是：应用程序先通过某种机制预先得到其地址对应在 NAT 上的外部地址，这样在负载中就直接填写 NAT 上的对外地址，而负载中的内容在经过 NAT 时就无需被修改，如 STUN 协议。

具备 STUN 类型穿越方式的应用，通常需要 NAT 设备具备非对称 NAT 能力，并且对于采用 TCP 方式连接的应用也不合适，因此目前大多数 P2P 流媒体应用需要 NAT 设备来支持外部地址端口开放能力，即 UPNP/PMP 方式。UPNP 方式可以通过发送控制信息在 NAT 网关上添加端口映射，来实现从外部访问 NAT 内部的目的，从而完成 NAT 的穿越。另外，对

于两个都处于 NAT 之后的节点，当它们需要通信时，互相无法知道对端的外部服务端口，通常某些应用软件通过端口试探或猜测的方式来完成 NAT 穿越，但是这样的方式效率低且成功率不能保证，所以通常需要中继服务器帮助建立连接，实现通信双方在中继服务器的信息中转。但是，在 NAT 被规模部署后，大多数通信双方都处于 NAT 后的情况会比较普遍，这时容易造成中继服务器的性能压力瓶颈，通过中继服务器的中转方式将不可行，这时既需要通过外部端口开放协议进行动态的外部端口开放，又需要具备非对称 NAT 的 NAT Server 服务，以增加安全性，这样通信双方可以直接建立连接，不再需要通过中继服务器的中转服务。

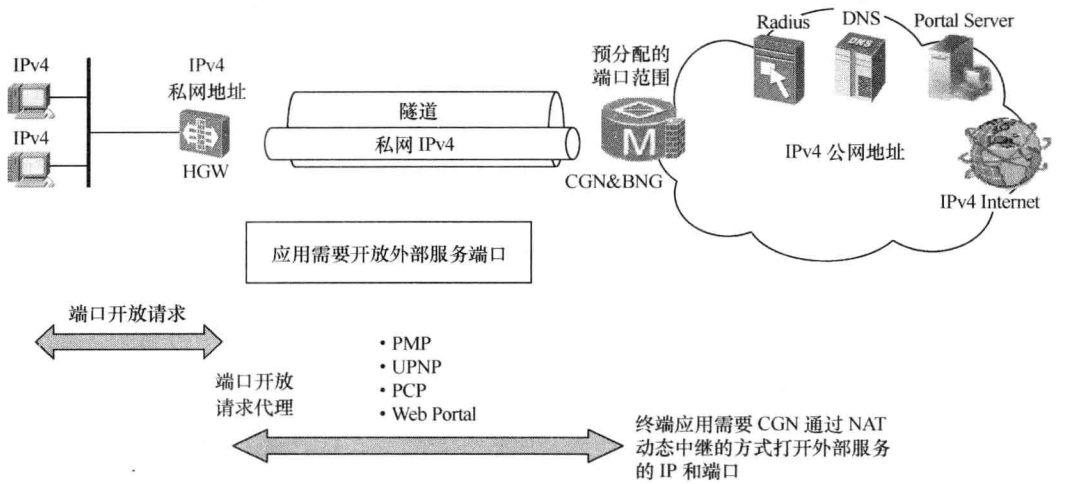


图 4-12　规模化部署 NAT 业务支持 NAT 穿越

电信级 NAT 需要支持大多数应用使用的动态端口开放能力，即 PMP、UPNP、PCP 等动态端口映射开放方式，结合 STUN 方式，实现规模化部署 NAT 的业务 NAT 穿越。这样的组合方式可实现 P2P 单/多通道业务的 NAT 支持，不用受限于特殊应用的 ALG 能力限制。类似的 PCP 和 Web Portal 方式也可以协商和管理外部端口，在某些应用下可满足特殊的要求。

11．NAT 业务级不中断可靠性

NAT 部署在网络核心位置时，对于可靠性和扩展性要求较高，为了达到电信级业务不中断，需要进行高可靠的冗余设计。单台 NAT 设备可通过多块 NAT 业务板卡实现板卡间的负荷分担和冗余备份，提高整机的设备处理性能和扩展性。对于多台设备的冗余和负荷分担，需要设备间的冗余备份机制来保证，NAT 设备间的备份机制如图 4-13 所示。

NAT 设备的可靠性备份机制基于扩展的 VRRP 功能，可以利用 VRRP 进行主备冗余备份的状态管理。VRRP 协议的扩展，用于实现设备间或板卡间多接入接口环境下的接口备份（华为公司的专利）。对 VRRP 协议的扩展主要在于传统的 VRRP 实现的是设备级的备份，而扩展后的 VRRP 实现的是接口级的备份，可以把一台设备的多个接口配置到同一备份组中。因此扩展后 VRRP 的备份粒度比传统 VRRP 更细，备份方式更加灵活。另外，传统的 VRRP 是一个三层协议，而扩展后的 VRRP 可以作为二层协议，用在二层用户接入侧场景下，因此无需配置接入接口的虚拟 IP 地址。NAT 利用扩展的 VRRP 协议，实现设备间或板卡间的备份。如果主接口、主接口所连接链路或主接口所在单板发生故障，将会把用户业务切换到备份接口。

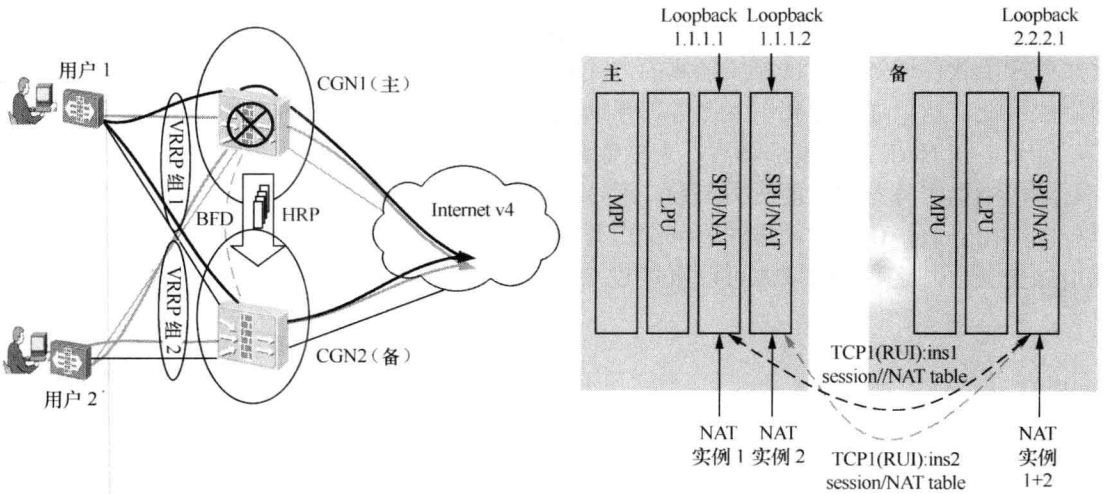


图 4-13　NAT 备份机制

主备设备间或板卡间的 NAT 用户信息需要实时同步，在主设备或板卡创建的用户管理控制表项信息需要同步到备用设备或板卡。一旦 VRRP 管理的接口或者接口链路发生故障，导致 VRRP 的主备状态发生切换，绑定在此 VRPP 上的用户立刻切换到备用设备上，并保证用户业务不中断，且没有分组丢失。设备间的信息备份机制采用 HRP 协议（华为公司的专利），实时同步状态信息。

对于管理 VRRP 组的主备状态检测，NAT 采用 VRRP 关联 BFD 技术，实现电信级的设备或链路故障检测。主备 VRRP 实体间 BFD 检测可以保证故障检测时间小于 50ms，这样主、备之间的平滑切换，用户感知不到，业务保持不中断。

系统内 VRRP 组数目有限，为了同时达到最大的管理灵活性，需要尽可能地细分管理的颗粒度，这样可以最大化地增加运营管理的灵活性。VRRP 管理组的备份关系可以关联相关的用户组，即 NAT 的管理实体、NAT 实例或与 NAT 实例关联的 BNG 的控制域，并将相应的 NAT 用户组绑定到 VRRP 管理组。一旦 VRRP 的备份组发生主备倒换，相关联的 NAT 用户组、NAT 实例或控制域用户切换到备用设备或板卡上，由备用设备处理后续的业务。

NAT 用户 session 实时在设备间热备同步，实时进行故障检测。备份关系支持 1∶1 或 N∶1 模式，以达到电信级 NAT 业务可靠性和扩展性要求。

4.4.2　DS-Lite 技术

DS-Lite（Dual-Stack Lite）技术的本质是在 IPv6 的网络中部署 IPv4-in-IPv6 隧道完成 IPv4 业务传输，而 IPv6 业务则直接通过 IPv6 网络传输。

“Lite”的中文意思是“轻量级的”，DS-Lite 的中文意思则为“轻量级的双栈”。所谓“轻量级”是相对双栈方案而言的。双栈方案要求感知 IP 的节点都具备双栈能力，且必须同时开启双栈能力。而 DS-Lite 方案则为“局部双栈”，换句话说，DS-Lite 技术可在 IPv6 网络中提供双栈孤岛互联。

该技术有一个很重要的特征，即：IPv6 的源只和 IPv6 的目的地址通信；同理，IPv4 的源只和 IPv4 的目的地址通信。这里没有不同地址族之间的互相转换，因而极大地简化了应用

层协议在载荷中携带 IP 地址的处理过程。所以，在这种网络架构下，IPv6 业务直接通过 IPv6 网络进行传输。那么 IPv4 业务呢？下面将重点阐述 IPv4 业务的传输。

1．基本概念介绍

（1）B4

B4 全称为 Base Bridging BroadBand Element。因为该名称中有 4 个 B，所以简称为 B4，该网元的功能如下。

① 具备双栈的能力，并在主机或者 CPE 设备上实现。其中，CPE 设备在运营商组网中为家庭网关。

② 向 AFTR 创建 IPv4-in-IPv6 的隧道。

（2）AFTR

全称为 DS-Lite Address Family Transition Router element。该网元终结了 IPv4-in-IPv6 隧道，并且实现了 NAT IPv4-IPv4 的功能。其中，AFTR 在运营商网络中为 CGN。该 CGN 可为独立 CGN 设备，也可以为插卡式 CGN（将 CGN 单板插在业务控制节点）。

更多详细信息可参考 RFC 6333。

DS-Lite 技术采用了以下两个大家熟知的技术来完成 IPv4 业务的传输。

① IPinIP（IPv4-in-IPv6）隧道：因 B4 和 AFTR 之间的网络为单栈的 IPv6 网络，所以家庭用户要访问 IPv4 业务需要通过 IPv4-in-IPv6 隧道进行传输。IPv4-in-IPv6 隧道从 B4 建立，终结于 AFTR。该隧道的封装方式为给 IPv4 报文直接封装 IPv6 的报文头。该封装报文就使用 IPv6 报文头在单栈的 IPv6 网络中进行转发。

② NAT（Network Address Translation）：家庭网络用户获得的 IPv4 地址为私网的 IPv4 地址，AFTR 终结 IPv4-in-IPv6 隧道之后，会将用户的私网地址转换为公网的 IPv4 地址。

2．逻辑组网图

（1）直连设备场景

图 4-14 显示的是直连设备场景。Host 承担了 B4 角色，AFTR 在 CGN 设备上实现。

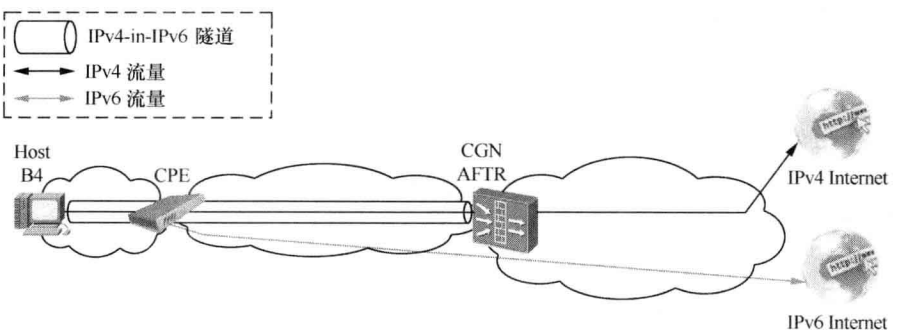


图 4-14　直连设备场景

该场景的特点如下。

① CPE 为桥接型家庭网关，其中 Host 为双栈主机。

② Host 和 CGN 之间为单栈的 IPv6 网络。

③ 主机和业务之间的网络必须具备 IPv4 和 IPv6 业务的传输能力。

该场景要求终端（如 Host）承担 B4 角色。因为不是所有类似主机的终端都可以支持 DS-Lite，而且运营商也不会提供类似主机的终端。

因此，从可部署的角度来讲，CPE 场景是被推荐的方案。

（2）CPE 场景

CPE 场景是指 CPE 作为 B4 的场景，如图 4-15 所示。

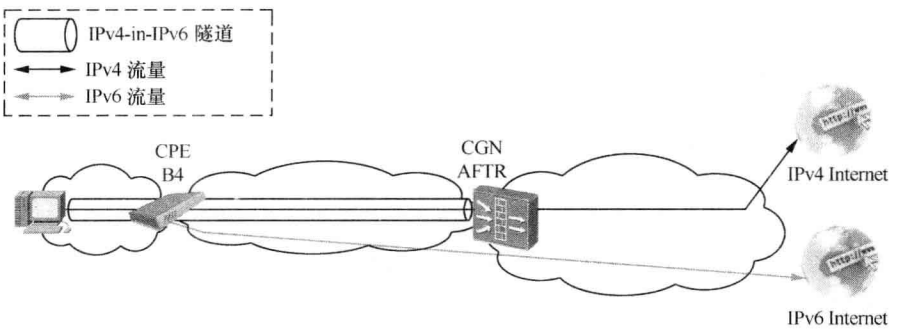


图 4-15　CPE 场景

CPE 的 WAN 口承担了 B4 角色，DS-Lite 技术不需要 CPE 的内部接口和 WAN 口之间具备 NAT 功能，而 CGN 承担了 AFTR（NAT IPv4-IPv4）角色。所以，该技术模型便避免了两级 NAT（一级在 CPE 设备上，一级在 CGN 设备上）的产生。

该场景的特点如下。

① CPE 为路由型家庭网关，其中家庭网络为双栈网络。

② CPE 和 CGN 之间为单栈的 IPv6 网络。

③ 家庭网络和业务之间的网络必须具备 IPv4 和 IPv6 业务的传输能力。

在该场景中，CPE 只能通过 DHCPv6 获悉一个 DNS recursive server 的 IPv6 地址，因为 CPE 和 CGN 之间的网络为单栈的 IPv6 网络，而 DHCPv6 协议也没有能力携带 IPv4 的 DNS 信息，只定义了获得 DNS recursive server 的 IPv6 选项（DNS Recursive Name Server option），所以 CPE 不能按照正常的方式获得 DNS recursive server 的 IPv4 地址。解决这一问题有如下两种方案。

① 方案一：采用 IPv6 报文存储 IPv4 DNS 内容，该技术为 IPv4-IPv6 DNS Proxy。例如，CPE 将家庭主机接收 DNS IPv4 的请求使用 IPv6 报文发送给运营商网络的 DNS Server。

② 方案二：如果 DNS recursive Server 的 IP 地址为 IPv4 地址，那么要通过 TR069、手工静态配置或者其他方式告知 CPE DNS 相关的 IPv4 地址，家庭主机会使用 IPv4 报文发送 IPv4 DNS 请求。一个 DNS 请求会在 AFTR 上创建一个绑定表。

值得注意的是，方案二中大量的 DNS 请求可能会直接对 AFTR NAT 表项的利用率产生影响，因此实际部署中推荐使用方案一。

不管是直连设备场景还是 CPE 场景，对 AFTR 网元的要求都是一样的，考虑到 CPE 场景实际可部署性较好，本书的关键技术将围绕着此场景进行介绍。

3．IPv4 地址分配

CPE 直接承担 DHCPv4 Server 将私网 IPv4 地址空间（例如 192.168.0.0/16）分配给家庭

用户，BRAS 只做 IPv6 地址分配。

如果家庭用户需要访问 IPv4 业务，那么会使用私网 IPv4 地址穿过 IPv4-in-IPv6 隧道抵达 CGN，CGN 会将私网 IPv4 地址转换为公网 IPv4 地址。

4．IPv6 地址分配

BRAS 只作 IPv6 地址信息管理，不作 IPv4 地址信息管理。

5．IPv4-in-IPv6 隧道

IPv4-in-IPv6 隧道建立在 B4 和 AFTR 之间。该隧道属于双向、无状态隧道。

在隧道建立之前，B4 必须获取 AFTR 的 IPv6 地址，AFTR 必须知道 B4 地址。B4 通过 DHCPv6 协议获取 AFTR 的名字，而 AFTR 通过 NAT 映射表知道 B4 地址。在实际的部署中，B4 地址为 CPE 的 WAN 口地址。

（1）AFTR 发现

为了建立 IPv4-in-IPv6 隧道，B4 需要获得 AFTR 的 IPv6 地址。B4 网元可通过 DHCPv6 选项获得 AFTR 名字，然后通过 DNS 业务将 AFTR 的名字解析为具体的 IPv6 地址。

（2）隧道建立

该隧道通过家庭网络中 IPv4 流量触发建立。

（3）隧道释放

在 PPPoE 接入模型中，该隧道会随着 PPPoE Session 的中断而释放。

在 IPoE 接入模型中，该隧道会随着对应邻居关系解除而释放。

（4）报文封装和解封装

报文封装和解封装的过程如图 4-16 所示。

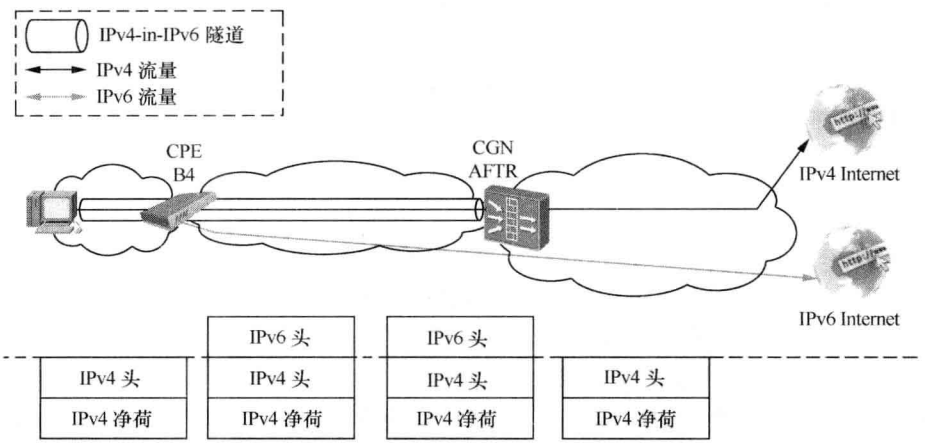


图 4-16　报文封装和解封装

假如 B4 下的用户需要访问 IPv4 业务，便发起一个 IPv4 报文。

① 封装：B4 将 IPv4 报文封装为 IPv6 报文，封装的方式为直接在 IPv4 报文上封装一个 IPv6 报文头（报文头的源地址为 B4 地址，目的地址为 AFTR 地址）。在 B4 和 AFTR 之间的 IPv6 网络中，该报文会使用 IPv6 报文头中的目的地址进行转发。

② 解封装：AFTR 收到该报文，会进行解封装。解封装的方式为将 IPv6 报文头剥离，

露出 IPv4 报文，然后 AFTR 会完成私网到公网的 NAT 转换。

IPv4 业务系统回应一个 IPv4 报文的过程如下。

① 封装：AFTR 收到该报文，根据 IPv4 报文中的三元组或者五元组信息找到对应的 NAT 映射表，进行公网到私网的 NAT 转换。然后，AFTR 可以根据私网信息（私网地址，私网端口号或者 ICMP Identifier）索引，检索到 B4 地址。然后，AFTR 会将 IPv4 报文封装为 IPv6 报文，封装的方式为直接在 IPv4 报文上封装一个 IPv6 报文头（其中报文头的源地址为 CGN 地址，目的地址为 B4 地址）。

② 解封装：B4 收到该报文后进行解封装，解封装的方式为将 IPv6 报文头剥离，露出 IPv4 报文。

（5）分片和报文重组

因为 PMTU 是 IPv6 传输的必选属性，所以 IPv6 报文在传输过程中可能会面临的分片和报文重组的处理。分片行为发生在隧道封装之后，报文重组发生在解封装之前。

6. NAT

NAT 绑定表扩展了入流量的源 IPv6 地址。入流量的源 IPv6 地址为 B4 的 IPv6 地址。在实际的部署中，该地址为 CPE WAN 口地址。不同的 CPE 会使用不同的 WAN 口地址，从而通过使用 IPv6 地址解决了家庭网络中主机的 IPv4 地址重复问题。

AFTR 的 NAT 功能继承了我们所知的 NAT 的所有特征，所以它同样也面临着 NAT 穿越、用户溯源等问题。

① NAT 穿越问题：同常规 NAT，例如可采用业内主流的方式 ALG。

② 用户溯源问题：可采取端口预分配的方式。端口预分配的原理就是预先评估和指定用户使用的外部端口号数量。端口号分配时，可以让用户使用的外部端口号分布在同一个公网 IPv4 地址范围之内。通过这种方式保证一个家庭的用户共享使用一个公网 IPv4 地址。当然，根据实际的使用情况，一个公网 IPv4 地址可能会被多个家庭使用。端口预分配的好处是用户信息具备明确的标识，容易溯源。

4.4.3 NAT64

NAT64 也是一种地址转换技术，用于 IPv6 终端访问 IPv4 服务，使得 IPv6 主机能够部分兼容原有 IPv4 业务。IETF 不准备开发 IPv4 终端访问 IPv6 业务的标准，主要考虑新技术向下兼容即可满足需要，并希望通过此策略推进 IPv6 发展。

NAT64 与 NAT44 的区别主要是其地址转换基于 IPv6 源地址和源端口，而不是 IPv4 源地址和源端口。NAT64 需要与 DNS64 配合，可以在现有的 DNS 上增加 DNS64 功能。NAT64 替代了原来的 NAT-PT（RFC 2766，NAT-PT 可支持 IPv4 终端访问 IPv6 服务）。NAT-PT 被废弃的原因在 RFC 4966 中做了说明。NAT-PT 的实现需要集成 DNS-ALG 的功能，而在 NAT 设备中集成应用层协议，在运营商环境部署会带来一系列难以克服的问题。

① 如何控制终端在访问 IPv4 时选择 NAT-PT 设备作为 DNS 的问题。

② 负载分担和可靠性问题

③ 应用层攻击，以及性能问题（NAT 的 CPU、内存资源远不如服务器）。

与 NAT44 类似，NAT64 主要适合 Client-Server 模式的应用，且数据报文的内容中不携

带地址和端口信息。引入 NAT64 是希望端到端部署纯 IPv6 网络，从而保持网络的简单性，但是目前应用层还没有准备好。

由于现阶段宽带互联网应用部署纯 IPv6 网络的基础尚不成熟，因此 NAT64 在目前固网中的应用相当有限，在移动互联网的应用要依赖运营商来推动产业链。由于移动互联网开放程度不如宽带互联网，推动应用开发商和服务提供商的难度要小。

NAT64 未来可能会以集中式部署为主。因为网络演进到 IPv6 占主流时，BNG 一般不会保留 IPv4 协议栈，而且 IPv4 业务量将比较少，分布式部署在成本上不合理。

4.5　过渡技术分析、对比

技术对比见表 4-4。

表 4-4　　3 种过渡技术的对比

	双栈	DS-Lite	NAT64
终端/主机	IPv4/IPv6 单栈或双栈	IPv4/IPv6 单栈或双栈	IPv6 单栈
CPE	支持桥接型 CPE，路由型 CPE 需要升级，最终网络路由型 CPE 占绝对主流	只支持路由型 CPE，CPE 需支持 DS-Lite 功能	不涉及
接入网	PPPoE 不涉及，IPoE 接入需要接入网支持 IPv6 特性	PPPoE 不涉及，IPoE 接入需要接入网支持 IPv6 特性	不涉及
SPOP	双栈	IPv6 单栈，一般集成 DS-Lite 转换网关功能	不涉及
骨干网	双栈	双栈或 IPv6	不涉及
特殊网关设备	无	DS-Lite 转换网关（AFTR），但单独部署将降低网络的可管理性	NAT64 网关
产业成熟度	好	一般，对应用层无要求，且无需 NAT444 两次转换	差，应用层支持少

4.6　过渡技术部署策略建议

运营商需要根据自己面临的问题，采用不同的演进部署方案。

对于面临竞争压力和政府政策部署 IPv6 的运营商，可选择双栈部署方案；对地址短缺很严重的运营商，可考虑部署 NAT 方案或部署双栈 + NAT 方案；而对于完全新建的网络，后续可考虑部署 DS-Lite（还要考虑相应的配套系统）。

当主流应用、终端都支持 IPv6 后，可考虑逐步关闭 IPv4 网络/协议栈，通过部署 NAT64 来解决少量的 IPv4 内容访问需求。

4.6.1　部署双栈，构建网络向 IPv6 迁移的基石

IPv6 部署应遵循“先核心、后边缘”的部署原则。首先骨干网/IGW 要支持双栈。根据现有网络部署技术和运维需求，可选择 Native 双栈和 MPLS 6PE/6vPE 两种路线。如果现网为 MPLS 网络，可选择 6PE/6vPE 路线，通过升级 PE 路由器来支持双栈，而无需对现有骨干

P 路由器做更改。

早期部署的业务需求主要来自于 VPN 专线。只需要升级或新建双栈 PE 路由器，便可满足要求支持 IPv6 的专线客户需求。

SPOP 是 IPv6 公众宽带业务部署的核心和难点。由于 PPPoE 业务无需改造接入网，HSI 早期部署建议首先从 PPPoE 宽带用户开始。通过升级或新建 BRAS 以及认证计费系统来支持个人用户双栈的接入。对于桥接式的 CPE，不需要进行更换，对于路由型 CPE，则需要升级为支持双栈。CPE 应可以通过软件升级支持后续演进。

最初引进 IPv6 阶段，也可以新建或升级少数 BRAS 双栈节点。其他区域 PPPoE 用户以 L2TP 接入双栈 BRAS。

对 VAS（Value Added Service）系统的升级改造应同时进行，以增加 IPv6 业务流量。

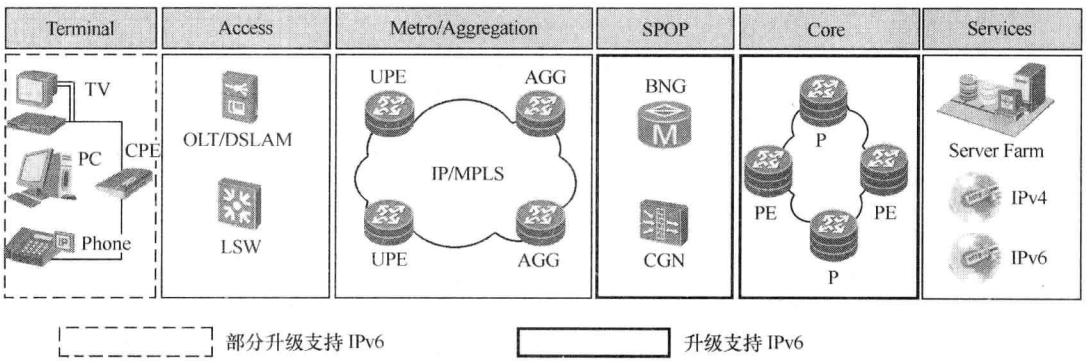


图 4-17　早期部署阶段，升级骨干网，改造 SPOP 和 CPE

4.6.2　部署 CGN，解决地址短缺，为业务平滑迁移提供保障

目前 IPv4 业务仍然占据主流，而 IPv4 地址短缺成为多数运营商需要面对的问题。NAT 是解决地址短缺的必然选择。一般地，引入 NAT 会导致网络流量模型变化，使得流量管理/用户业务识别的难度增加，甚至使得某些业务受损。

BNG/CGN 融合方案，通过 BNG 集成 DS-Lite 或 L2 Aware NAT。业务只需一次 NAT，而且这种分布式 NAT 结构降低了对设备的性能要求，容易隔离故障和定位，运营商也无需管理终端的 IPv4 地址，无需部署 NAT Log，提高了用户对业务的感知。缺点在于：更换终端的成本较高（现有 CPE 终究要更换，以支持 IPv6）；依赖于 BRAS 厂商的支持；依赖于使用 IP 地址识别用户的系统，如流量检测、流量管理等系统都将不能正常工作。

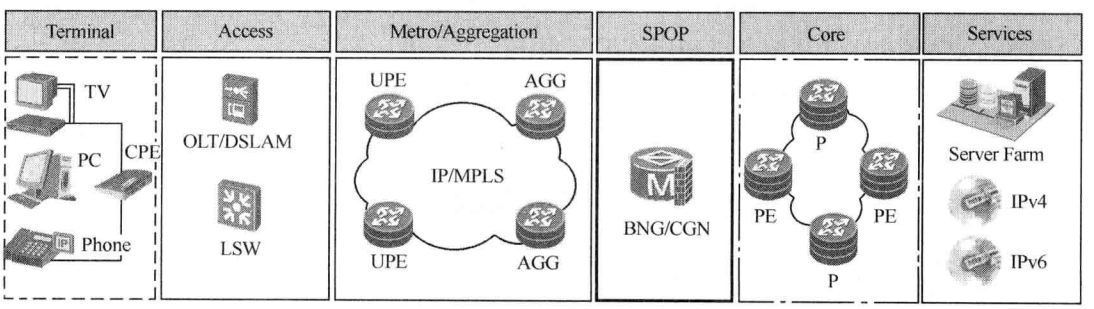


图 4-18　部署 BNG/CGN 应对地址短缺

集中式 NAT 目前已经被广泛部署。在网络出口放置集中式 NAT，对私网地址用户的业务进行重定向，转到 NAT 设备上进行地址转换。独立式 NAT 部署的优点在于：设备可获得性好，原有网络和业务系统几乎不需要做更改。缺点在于：集中式设备故障影响面广，且故障定位较困难；需要建设溯源系统；对 IPv6 部署没有明显促进。

在双栈网络中，解决地址短缺依然可以采用集中式 NAT 方式，如图 4-19 所示。

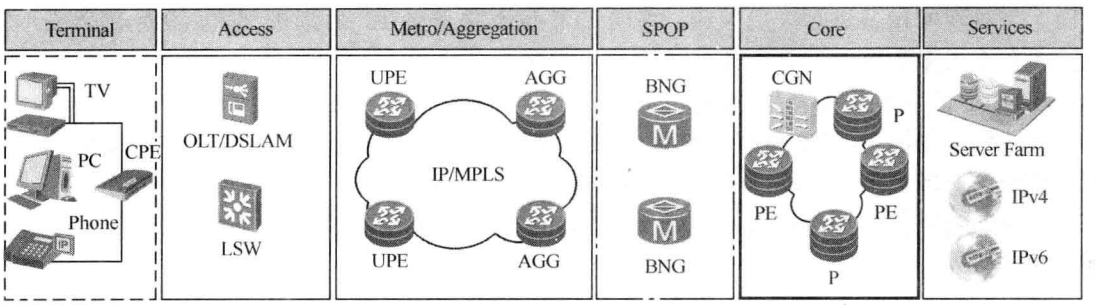


图 4-19　部署集中式 NAT 应对地址短缺

业界已经达成共识：DS-Lite 是同时解决地址短缺与加速 IPv6 发展的最佳方案。其相关标准已经成熟，被多个主流运营商选中作为演进路线，成为 Native 双栈方案之外唯一的主流演进技术。终端和设备供应商已有支持 DS-Lite 的商用产品发布。

两种主流的演进路线是可以兼容的。在初期部署阶段可采用 Native 双栈，同时进行 DS-Lite 试验（DS-Lite 只使用网络的 IPv6 协议栈），在条件具备的情况下，可逐步切换到 DS-Lite 方案。

4.6.3　向纯 IPv6 网络过渡

向纯 IPv6 网络过渡主要解决 2 个问题，即：接入网支持 IPv6，以及部署 NAT64。

随着接入方式逐步向 IPoE 迁移，对接入网支持 IPv6 业务的感知需求越来越迫切。在 IPoE 接入场景下，IPv6 协议报文一般为特定多播报文，而且原有 DHCP option、DHCP Snooping、IP/MAC 绑定、多播 Snooping/proxy 等安全特性都需要针对 IPv6 进行开发。此外，在 IPv6 协议下，ICMPv6 为接入层的新攻击点，Neighbor Discovery Snooping 成为接入层设备必备的安全特性。

随着 IPv6 网络部署的推进，对成本、功耗、海量且简单部署的需求将导致纯 IPv6 终端爆炸性增长。部分终端可能存在访问 IPv4 业务的需求，通过部署支持 NAT64 的电信级 NAT，可使得 IPv6 终端以可管理的方式访问 IPv4 网络和业务。

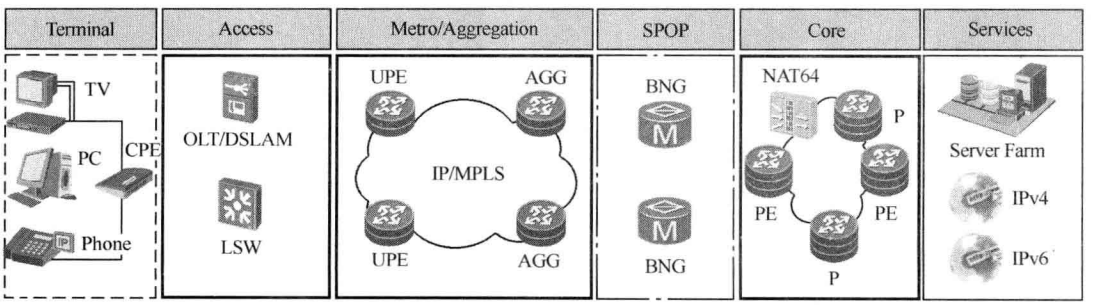


图 4-20　升级接入网，部署 NAT64 和纯 IPv6

4.6.4 全球运营商 IPv6 演进方案选择

根据调查，全球大部分运营商倾向于选择双栈方案来部署 IPv6，并采用 NAT 解决地址短缺问题。也有部分运营商有意选择 DS-Lite 作为演进方案，但目前尚无商用部署案例。而极少数采用 6to4、6RD 做早期实验部署的运营商，也逐步减少了对 6RD 的投入。

选择双栈 + NAT 的运营商主要的考虑因素是：地址短缺已非常严重，急需解决，而选择双栈无需对现网设备进行大量改造，终端 CPE 也不急于更换，可渐进部署，初始部署成本较低，适合绝大多数对资本支出比较敏感的运营商。

DS-Lite 的主要推动者是 FT 和 Comcast，这两家公司在 DS-Lite 领域研究投入较多，拥有专利和标准影响力。FT 选择三层到边缘的网络架构，所以希望采用纯 IPv6 以简化管理。而且其现网已经部署的 CPE 具备软件升级支持 DS-Lite 的能力，成本增加相对较小。

部分运营商正计划对双栈 + NAT 和 DS-Lite 进行试验，在试商用中评估选择最合适的演进方案。

第 5 章
运营商过渡技术部署

5.1 运营商网络演进策略总体分析

IPv4 地址资源即将枯竭，IPv6 是解决地址问题的根本手段，所以运营商网络从 IPv4 向 IPv6 演进是一个必然。运营商需要考虑的是如何在地址枯竭前完成从 IPv4 向 IPv6 的平滑演进。

当前 IPv6 端到端产业链的状况呈两头弱中间强的纺锤状，如图 5-1 所示。所谓两头弱，是指固定、移动终端以及业务系统对于 IPv6 支持度低，升级困难；中间强是指中间的城域、骨干、承载等网络部分对 IPv6 的支持程度较高，当前主流电信制造商已研发了大量 IPv6 产品，现网大部分产品可以支持升级至 IPv4/IPv6 双栈。

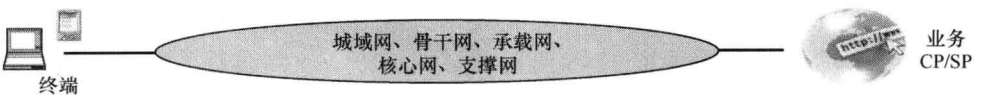


图 5-1 IPv6 端到端产业链

（1）终端当前不成熟

① 家庭网关：现网路由 HG 均不支持 IPv6，家庭网络升级风险大。

② STB：STB 机顶盒需通过软件升级才能支持 IPv6。

③ 终端升级/替换成本大。

（2）业务系统当前不成熟

① 全球网站排行榜前 100 万个网站只有不到 1%的支持 IPv6。

② 业务系统改造成本大。

（3）网络部分当前成熟，可以改造

① 从骨干网到城域网，主流厂家网络设备已支持 IPv6，接入网改造有压力，现网多数设备可升级。

② IPv4 向 IPv6 过渡技术走向成熟，过渡方案争议较大。

② GSN/PDSN：主流厂家 GGSN 等核心网设备均已支持 IPv4/IPv6 双栈。

③ 支撑网支持双栈条件具备。

由如上分析可以得出，当前运营商网络层基本具备向 IPv6 网络的演进条件。对于运营商来说，IPv6 是解决目前 IPv4 地址短缺问题的根本方案，但是需要一个逐步演进的过程。在向 IPv6 网络迁移的过程中，必须同时考虑下面的问题。

① 如何继续维持和提供目前的 IPv4 业务？

由于 IPv4 和 IPv6 报文头不兼容，因此，现网中的基于 IPv4 开发的设备和主机必须升级才能支持 IPv6。而目前大多数互联网应用还是基于 IPv4 的，支持 IPv6 的应用还非常少。因此，在迁移过程中必须继续维持目前的 IPv4 业务，否则就意味着用户的丢失。

② 如何基于现有 IPv4 网络，升级和构建 IPv6 基础网络？

由于现有 IPv4 网络设备和主机数量庞大，升级替换成本巨大，升级替换周期很长，还需考虑网络设备和主机的协调。因此需要考虑在为多种业务提供 IPv6 通路的同时，保证 IPv6 网络可靠性、可扩展性、可管理性、QoS、安全、移动、多播等能力。

③ 如何将 IPv4 网络平滑迁移到 IPv6 网络？

IPv6 的引入是一个渐进过程，很长时期内 IPv4 和 IPv6 网络将会共存。对运营商来说，一方面需要尽快引入、部署 IPv6 网络，解决 IPv4、IPv6 网络共存和互通问题；另一方面也要解决现有业务持续发展所要面临的 IPv4 公有地址不足的问题。因此，需要考虑保证 IPv4 网络到 IPv6 网络平滑迁移的措施。

5.2 网络部署

5.2.1 网络架构和改造关键点分析

在从 IPv4 网络向 IPv6 网络演进方案的选择上，鉴于目前 IPv6 端到端产业链没有完全成熟，应该统筹考虑。既需要通过 NAT 技术解决短期内 IPv4 地址不足对业务发展的限制，也要长远考虑网络最终向 IPv6 演进的必然趋势。控制演进节奏，节省演进成本，循序渐进，逐步推进，既要充分保护 IPv4 现网设备投资，同时也要确保现有业务体验不变。

过渡机制基本可以分为 3 类，即：双栈（Dual Stack），隧道（Tunnel），翻译（Translation）。3 种基本的过渡技术中，双栈是最为简单和易被接受的过渡方式，隧道技术和翻译技术是特定场合的必要技术。

1. 双栈技术

所谓双栈（Dual Stack），就是在一个系统中可以同时支持 IPv6、IPv4 两个可以并行工作的协议栈，是最直接、最简单引入 IPv6 并实现 IPv4 与 IPv6 共存的方案。网络节点同时运行 IPv4 和 IPv6 两种协议，网络在逻辑上成为 IPv4 及 IPv6 并行的两个网络，支持向 IPv6 的平滑过渡。当前，双栈技术是一个比较成熟的过渡技术，对网络的影响最小，而且没有 ALG 和 MTU 问题。缺点是需要全网设备升级，投资大。双栈方案虽然不能解决 IPv4 地址短缺的问题。却是向 IPv6 发展演进的关键阶段。在过渡技术中，一般建议优先使用双栈，同时配合使用其他过渡技术。

双栈技术是 IPv6 演进的基础，近来提出了针对实际部署的多种演进技术的组合方案——针对不同阶段、不同场景下的双栈和隧道、转换结合技术。

2．隧道

隧道是用来连接网络中出现的 IPv4/IPv6 孤岛的有效方法，一般用于网络中 IPv4/IPv6 单栈流量占优的场景。在 IPv6 网络演进初期，6in4 的隧道可以用来连接 IPv6 的孤岛。当 IPv6 网络演进到一定程度的时候，IPv4 将成为孤岛，4in6 隧道则用来连接 IPv4 的孤岛。使用隧道技术可以较好地利用现网存量设备，而且没有 ALG 问题。针对不同的应用场景，当前常用的隧道包括：6to4 隧道、L2TP 隧道、6PE/6VPE、6RD、DS-Lite。

隧道技术用于实现分布于 IPv4 网络中孤立的 IPv6 网络之间的互连，或者分布在 IPv6 网络中的 IPv4 岛屿互连。隧道技术只需要边界节点实现双栈，并通过隧道将一个地址族的数据穿越到另一个地址族网络。

在 IPv6 发展初期，隧道技术便于实现 IPv6 孤岛间的互通，便于增量部署 IPv6，不需整网升级，从而逐步扩大了 IPv6 的实现范围，因而是过渡初期最易于采用的技术。同样，在过渡后期，分散的 IPv4 孤岛间也可以使用隧道连接。

隧道的缺点在于：双层 IP 头增加了网络开销，隧道端点需要额外处理扩展性和可靠性，以及可能的 MTU 问题。

常用的隧道特性和场景见表 5-1。

表 5-1　　常用隧道技术对比

隧道类型	技术特征	适合的场景
手工隧道	采用 IP-in-IP 或 GRE 封装	隧道为手工配置，实现简单，网络设备广泛支持，但不适合规模部署
自动隧道	采用 IPv6-in-IP 方式，利用 IPv6 嵌入 IPv4 地址，实现无状态自动隧道封装	只能用于 IPv6-in-IPv4 隧道，通过嵌入 IPv4 地址实现自动隧道，依赖 IPv4 拓扑，适合 IPv6 过渡初期使用，适用于主机，需要常用操作系统支持。其中，6to4 需要公网 IPv4 地址，主要用于 IPv6 孤岛间互连，与 Native IPv6 通信需要 6to4 Relay Router
MPLS 隧道	6PE/6VPEIPv6-in-MPLS	隧道转发性能好，适用于网络核心，要求有 MPLS 基础网络

除此之外，6RD 延续了 6to4 无状态隧道方式。由于 6RD 采用运营商前缀代替 6to4 使用的 Well-known 前缀 2002/16，从而使 IPv6 前缀可以在 Native IPv6 网络中发布，优化了 Native IP 网络访问 6to4 孤岛的路由问题。

3．转换

当前常用的地址转换技术包含有 NAT44 和 NAT64。

NAT44 是通过采用公/私网地址转换的技术来保守地延长 IPv4 使用的方法，是一种较为成熟的过渡技术，可以很好地解决运营商地址不足的问题。但该技术存在 ALG 的问题、扩展性问题，一般要尽量避免两级以上的 NAT。

NAT64 是通过 IPv6/IPv4 地址转换技术来解决 IPv6-only 和 IPv4-only 通信的唯一方法，一般用于 IPv6 演技技术的末期。

翻译技术用于 IPv6-only 网络和 IPv4-only 网络的互通。翻译设备位于两个网络的边界，翻译时需要将 IP 报文头的各对应字段强制转换。此外还需要将报文体中携带的地址实现转换。后者称作 ALG（Application Level Gateway）。

翻译技术种类较多，表 5-2 列出了主要翻译机制的技术特征和应用场景。

表 5-2　　常用翻译技术对比

层次	翻译技术	技术特征	适合的场景
网络层	SIIT	定义了用特定地址格式实现地址的转换：IPv4 映射地址 0：：ffff：a.b.c.d 和 IPv4 翻译地址 0：：ffff：0：a.b.c.d	无状态翻译，面临 IPv4 地址不足的问题，只适用于 IPv6 Network 到 IPv4 Internet 的场合
	NATPT	通过NAT技术实现IPv4地址＋端口与 IPv6 地址的对应关系	有状态翻译，置于路由器或防火墙内，目前最成熟，相比应用层效率更高
	BIS	主机中实现的 NAT-PT	适用于单栈主机
传输层	TRT	传输层的翻译	适用于路由器
应用层	SOCKS64	使用 SOCKS 协议，通过 SOCKS 代理服务器实现地址翻译	需要升级主机软件并部署专门的 SOCKS 服务器，适用于特定应用场合
	BIA	主机中实现的 SOCKS64	主机翻译技术，应用于双栈主机的传统应用程序
	ALG	应用层内实现地址转换	性能较低，通常与网络层翻译技术配合使用，用于报文内容的翻译

然而，由于翻译技术本身存在的问题，目前只有 NAT-PT（Network Address Translation-Protocol Translation）有产品实现，可部署，其他方法鲜有被关注。网络层上的翻译技术因处理层次多，性能很低。网络层的翻译技术主要面临的是 ALG 问题，部分协议的报文体中带有 IP 地址，所以翻译设备必须识别特定的应用层协议才能够实现 ALG，并可能对新出现的应用形成阻碍。

IETF 已经注意到翻译技术面临的问题，在 RFC 4966 中有问题的详细描述，包括 DNS-ALG 问题、NAT 问题、ALG 及扩展性等。在此基础上，IETF 的 Behave 工作组重新思考翻译技术，并提出改进。

有状态翻译采用 NAT64＋DNS64 简化和替代 NAT-PT。NAT64 只允许 IPv6 侧单向发起连接，并且不再实现 NAT-ALG，后者功能由单独的 DNS64 服务器来实现。

无状态翻译吸收了 SIIT 和 IVI 的思路，使用运营商前缀。IPv6 地址内嵌 IPv4 地址，并且可以自映射，更易于地址管理。

无状态翻译重新定义了 IPv4 和 IPv6 分组头翻译的对应关系，以及地址的翻译格式；定义了 IPv4/IPv6 翻译时利于业务互通的 NAT 行为方式。

IPv4 向 IPv6 迁移有多种方案，可分别用于不同的阶段和不同的场景。在选择 IPv6 网络过渡技术时要综合考虑以下因素。

① 保护现网投资，成本合理，网络改造难度较低。

② 有业务不受损，用户体验好。

③ 能够推动、鼓励 IPv6 的发展。

5.2.2　骨干网

在现有的 IPv4 网络架构中部署 IPv6，并解决 IPv4/IPv6 网络共存和设备互通的最自然的方式是在终端和网络节点上既安装 IPv4 又安装 IPv6 的协议栈，即双栈，从而实现 IPv4 或 IPv6 节点间的信息互通。根据网络业务的不同，常用的骨干网 IPv6 过渡技术有 6PE/6VPE、双栈。

1．双栈

双栈是运营商网络向 IPv6 演进的基础。由于骨干网占用 IPv4 的地址较少，主要职能为路由和承载 IP 数据流，所以不需要采用 NAT 等技术缓解 IPv4 地址不足的问题。重点需要考虑的是如何引入 IPv6，实现网络的平滑演进。

IPv4/IPv6 双栈方案是最直接、最简单引入 IPv6 并实现 IPv4 与 IPv6 共存的方案。网络节点同时运行 IPv4 和 IPv6 两种协议，网络在逻辑上成为 IPv4 及 IPv6 并行的两个网络，支持向 IPv6 的平滑过渡。

对于没有启用 MPLS 的 IPv4 骨干网络，采用分平面逐步升级 IPv4/IPv6 双栈的策略。首先在 P 设备上开启 IPv4/IPv6 双栈路由功能，最终过渡到全网双栈网络甚至是 IPv6 单栈 MPLS 骨干网络，如图 5-2 所示。

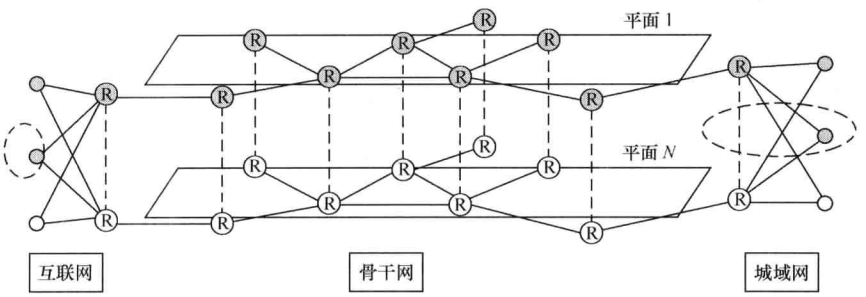


图 5-2　骨干网分平面升级双栈示意图

2．6PE/6vPE

骨干网设备位于网络核心。现网老设备存量较多，全部升级替换代价较大。所以，骨干网除了要考虑如何平滑引入 IPv6 业务，还需要充分保护运营商 IPv4 现网设备投资。

在骨干网络部署了 IPv4 MPLS 的情况下，可以在骨干网络 PE 设备升级支持 6PE 或 6vPE 特性，以使 IPv6 的流量通过 MPLS 隧道连通。这种方案可以保持骨干网络核心 P 节点仍为 IPv4，在骨干网仅部署 IPv4 的路由协议，如图 5-3 所示。

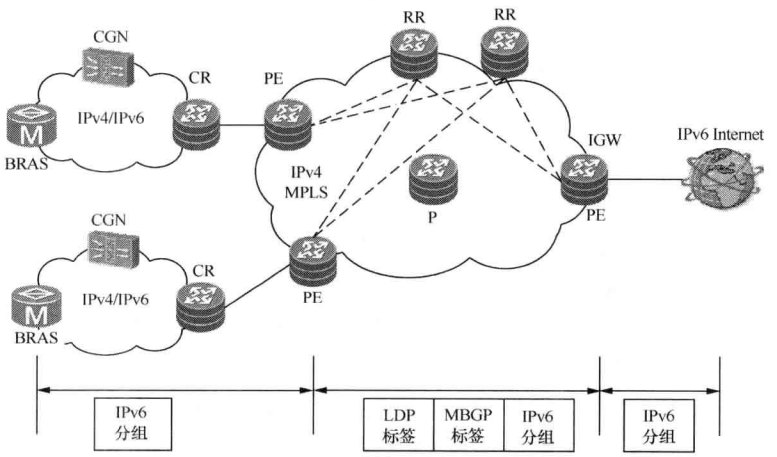


图 5-3　IPv6 通过 MPLS 网络承载方案

6PE/6vPE 方案的基本原则是：重用现有的 IPv4 路由和 MPLS 结构，以及 PE 节点升级支持 6PE、6vPE。

使用 6PE/6vPE 作为 IPv6 网络间连通的隧道的主要优势在于以下几点。

① 所有配置在 PE 上完成，用户网络感知不到 IPv4 网络的存在。

② 能够很好地利用 ISP 现有的 MPLS 网络资源，对运行商网络改造不大。

③ E-CE 之间链路可以使用任何类型，没有特殊要求。

④ 6PE/6VPE 设备可以同时为用户提供 IPv6，IPv6 VPN 和 IPv4 VPN 等多种业务。

（1）6PE 方案

6PE（IPv6 Provider Edge）是在目前的 IPv4 网络中利用 MPLS 隧道技术为不同地区的 IPv6 网络提供连通服务，能够很好地利用现有的 MPLS 网络资源。6PE 方式是在 ISP 的 PE 设备上实现 IPv4/IPv6 双协议栈，利用 MP-BGP 在 PE 间发布 IPv6 路由信息，PE 之间通过 Tunnel/LSP 来透明传输 IPv6 报文。

该方案需要将原 IPv4 骨干网的 PE（包括 IGW）升级为 IPv4/IPv6 双栈，并部署 6PE 特性。

部署该方案时，核心节点 P 保持不变，仍为 IPv4 路由器。PE、P 之间利用 IPv4 路由协议交换路由，建立 MPLS 隧道。建立 MPLS 隧道的方式可保持原来的 IPv4 LDP 或 RSVP TE 的方式，如图 5-4 所示。

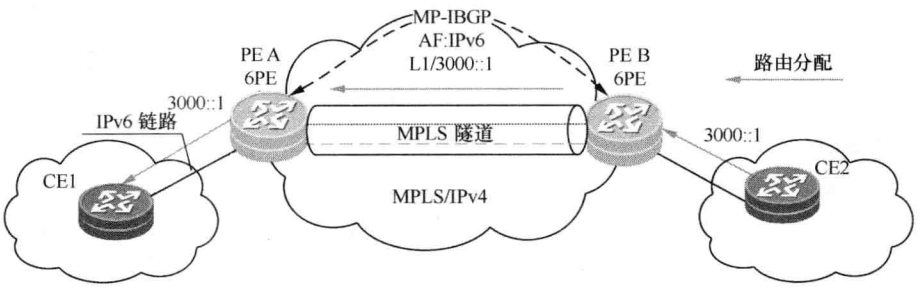


图 5-4　6PE 方案示意图

PE 和 CE 之间利用 IPv6 协议中的 IGP、EBGP 和静态路由等方式交换 IPv6 路由。6PE 不是 VPN 技术，只是通过 IPv4 骨干网络来连接 IPv6 网络，使 IPv6 网络通过 IPv4 骨干互通。如果 PE 具备 MPLS 功能，但 P 设备只提供纯 IP 功能而不具备 MPLS 功能，则可使用 GRE 隧道。

（2）6vPE 方案

6vPE 方案就是在 IPv4 MPLS 骨干网中建立 IPv6 BGP/MPLS VPN，构建一个 IPv6 VPN 的网络。6vPE 原理与 BGP/MPLS IPv4 VPN 相同，通过 MP-BGP 在 PE 间发布 VPN-IPv6 路由信息，PE 之间通过 Tunnel/LSP 来透明传输私网 IPv6 报文，如图 5-5 所示。

部署 6vPE 时，需将 IPv6 VPN 划分为管理 VPN、个人业务 VPN，以及企业专线 VPN 等类型，从而实现业务的隔离。

该方案需要将原 IPv4 骨干网的 PE（包括 IGW）升级为 IPv4/IPv6 双栈。如果原来的 VPN 中部署了 RR，也需将 RR 升级为 IPv4/IPv6 双栈，用于反射 IPv6 路由。在 PE 上部署 6vPE 特性，在 PE 之间建立 IPv6VPN 隧道，连接各 IPv6 Site 与 IPv6 Internet，以使 IPv6 VPN 报文通过 MPLS VPN 隧道穿越 MPLS 骨干网络。

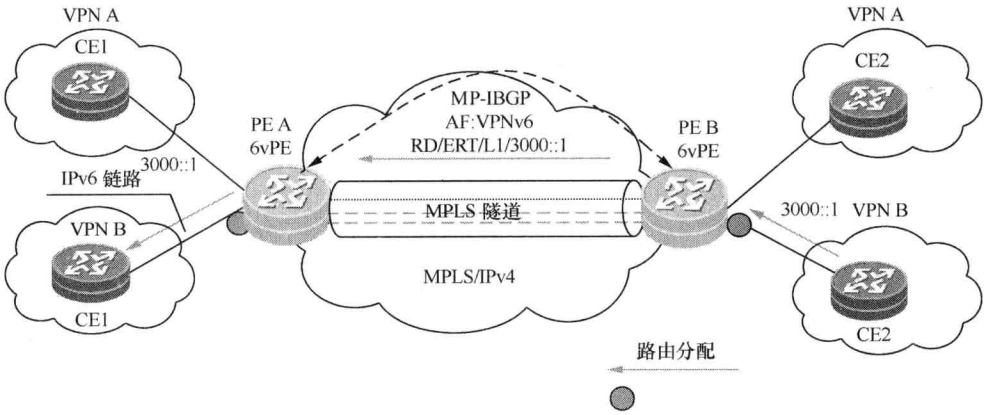


图 5-5　6vPE 方案示意图

与 6PE 不同，IPv6 VPN PE 间交互路由的时候，带上了类似于 IPv4 VPN 的 RD 和 ERT（VRP 使用 VPN-Target 属性控制 VPN 路由的收发，用于 VPN 互访控制，从而实现多种 VPN 组网拓扑）。PE 间传递的标签 L1 是 VPN 标签，可以标识出 CE 接入的 VPN。

核心节点 P 保持不变，仍为 IPv4 路由器。PE、P 之间利用 IPv4 路由协议交换路由，建立 MPLS 隧道。建立 MPLS 隧道的方式可保持原来的 IPv4 LDP 或 RSVP TE 的方式。

PE 和 CE 间的路由交换可以采用静态路由、OSPFv3 多实例、IS-IS 多实例，也可以采用 BGP4 + 。

PE 之间需要创建 Tunnel 来透明传输 IPv6 报文。PE 和 P、其他 PE 之间利用 IPv4 路由协议交换路由，建立隧道。一般使用 LSP 或 MPLS TE 隧道作为 VPN 骨干网隧道。如果 PE 具备 MPLS 功能，但 P 设备只提供纯 IP 功能而不具备 MPLS 功能，则可使用 GRE。

（3）6PE/6vPE 总结

初期采用 6PE/6vPE 隧道技术可以很好地解决现网老旧设备数量众多，需要兼顾投资效益的问题。在 PE 边缘节点支持 IPv4/IPv6 双栈接入，P 节点保持 IPv4 单栈，后续随着设备的升级替换逐步将 P 设备开启 IPv4/IPv6 双栈，最终过渡到全网双栈网络甚至 IPv6 单栈网络。

初期，骨干网可考虑采用 6PE 技术，实现单栈 IPv4 的骨干网将各省级 IPv6 孤岛网络连接起来。在 PE 边缘节点，支持 IPv4/IPv6 双栈接入，骨干网中 P 节点仍然保持 IPv4 单栈。

5.2.3　城域网

城域网的过渡技术种类较为繁多，但 IPv4/IPv6 双栈方案同样是演进的基础，在网络节点同时运行 IPv4 和 IPv6 两种协议，网络在逻辑上成为 IPv4 及 IPv6 并行的两个网络，支持向 IPv6 的平滑过渡。

与骨干网不一样的是，城域网中接入了大量的宽带上网用户，有着 IPv4 地址紧缺的问题。双栈方案虽然不能解决 IPv4 地址短缺问题，却是向 IPv6 发展演进的关键阶段。从业务发展的角度看，可以通过双栈部署，积累运营商 IPv6 网络运营及业务拓展经验，培养终端用户 IPv6 业务使用习惯。一般，各种过渡技术会结合双栈一起应用。常见的城域网过渡技术有 6RD、双栈、L2TP、NAT444、DS-Lite。

1. NAT444 + 双栈

由于双栈设备还需要继续分配 IPv4 地址，因此在 IPv6 的迁移完成前，仅部署双栈并不能解决现网 IPv4 地址紧缺问题。

当 IPv4 公有地址不足时，采用 IPv4 私有地址给用户提供接入服务成为一个无法避免的选择，网络地址转换 NAT（Network Address Translation）的使用也就成为必然。NAT 通过使用少量的公有 IPv4 地址映射大量的用户私有 IPv4 地址的方式，可以在一定程度上缓解 IPv4 地址空间枯竭的压力，是当前已经广泛应用的技术。

双栈和 NAT 技术的结合，可以在解决 IPv4 地址短缺的同时，支持已有网络平滑升级到 IPv6，是当前运营商部署 IPv6 的主流选择，如图 5-6 所示。

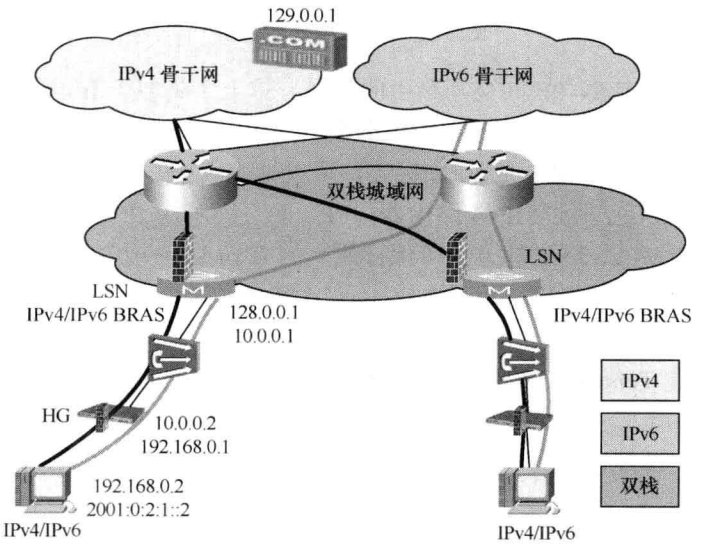


图 5-6　双栈+NAT 技术

（1）NAT444 方案是两级 NAT44 的结合

第一级 NAT44 在家庭网关上实现用户侧 IPv4 私网地址到运营商分配的 IPv4 私网地址的 NAT 翻译，第二级 NAT44 在运营商网络中的 LSN 上实现运营商分配的 IPv4 私网地址到 IPv4 公网地址的 NAT 翻译，从而提高 IPv4 公网地址复用率，缓解 IPv4 地址枯竭。

NAT44 已经应用多年，技术成熟度高。目前市场上现有的家庭网关基本上都已支持 NAT44 功能，运营商只需要在城域网中增加 LSN 设备就可以实现，无需进行较大规模的设备升级替换，部署成本相对较低。

（2）NAT444 对 CPE 的要求

① HG 路由模式：需要做 NAT，将 CPE 域内的私有 IPv4 地址翻译成 CGN 域内的私有 IPv4 地址。大多数路由模式 HG 当前已经具备此功能。

② HG 桥接模式：NAT444 对桥接型的 HG 无要求。

（3）NAT444 大规模部署存在的问题

但 NAT444 方案大规模部署也存在一些问题，最突出的两个是地址溯源及业务穿越。

地址溯源分为离线地址溯源与实时地址溯源两种。离线地址溯源需要记录大量 NAT 日

志，同时提供方便的查询工具。离线溯源已经有小规模的商用部署经验可以借鉴。如果需要全面部署，运营商需要制定统一的日志格式，以实现 NAT 日志的统一管理。实时地址溯源则更加复杂，目前没有成熟的方案，更没有成功的商用部署，需要运营商与设备商定义 LSN 设备、AAA 以及实时日志查询系统之间的标准接口及交互信息，同时充分评估 LSN 设备对实时查询量的支撑性能。

对于业务穿越，LSN 设备上需要开发与业务相对应的 ALG（Application Layer Gateway，应用层网关）以实现业务的穿越。但不断出现的新业务可能会需要 LSN 不断开发新 ALG，从而造成网络设备的频繁升级，影响网络稳定性。目前主流的看法认为未来的业务提供商应该向现在的 MSN、QQ 一样，通过业务系统升级支持 NAT 穿越。但这一点还需要产业链沟通协调，相互理解配合。

2．DS-Lite（Dual Stack Lite）：轻量级双栈

DS-Lite 是轻量级双栈方案，网络部署上只需要家庭网关（B4）和运营商网络中新增部署的 AFTR（Address Family Translation Router）为双栈，其他网络节点只需要支持 IPv6 即可。运营商为家庭网关（B4）分配私网 IPv4 地址及公网 IPv6 地址。用户访问 IPv6 业务可通过家庭网关及 IPv6 城域网直接到达。对于 IPv4 业务，则通过家庭网关（B4）与 AFTR 之间建立 IPv4-in-IPv6 的隧道，在 AFTR 上实现隧道的解封装，并将 IPv4 私有地址转化为 IPv4 公网地址，发送到 IPv4 业务系统。通过 DS-Lite 方案部署，运营商可以通过一个 IPv6 网络同时提供 IPv4 与 IPv6 服务，如图 5-7 所示。

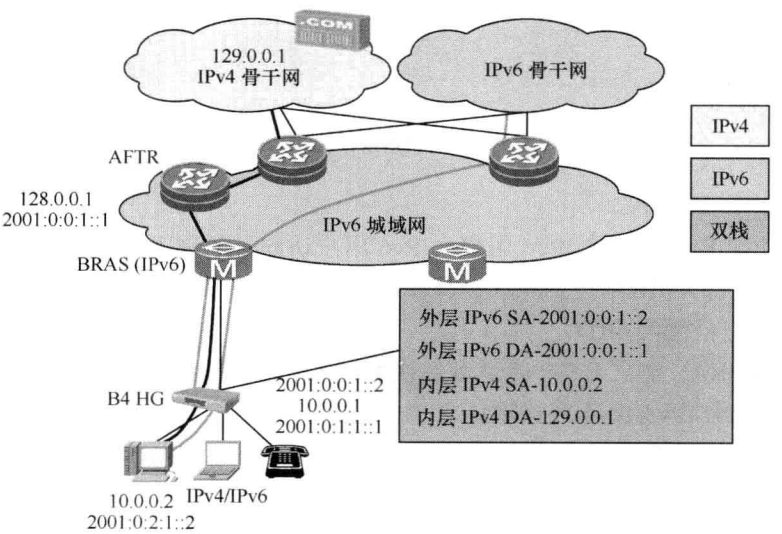


图 5-7　轻量级双栈

与 NAT444 方案相比较，DS-Lite 对于 IPv6 的演进态度更积极，更能促进 IPv6 产业链的发展。但 DS-Lite 方案为了节省 IPv4 公网地址也存在 NAT 技术的应用，所以 NAT444 方案面临的地址溯源及业务穿越问题在 DS-Lite 方案中同样存在，也需要类似的解决方案以满足大规模商用部署的要求。

DS-Lite 方案的前提是城域网为 IPv6 网络。DS-Lite 的场景其实是 4over6 的隧道。

DS-Lite 的应用场景包括以下几部分。

① 终端为 IPv4 终端。

② 城域网为 IPv6 单栈网络。

③ 目的为访问 IPv4 骨干网。

3. L2TP + 双栈

从保护投资的角度看，现网城域网设备 60%以上可以通过软件升级支持双栈，不能升级支持双栈的设备也可以通过 L2TP 隧道技术接入双栈，最大限度地使用现网设备，节省新建设备投资。从技术可行性角度看，双栈是标准、设备、商用部署实践最成熟的过渡技术，如图 5-8 所示。

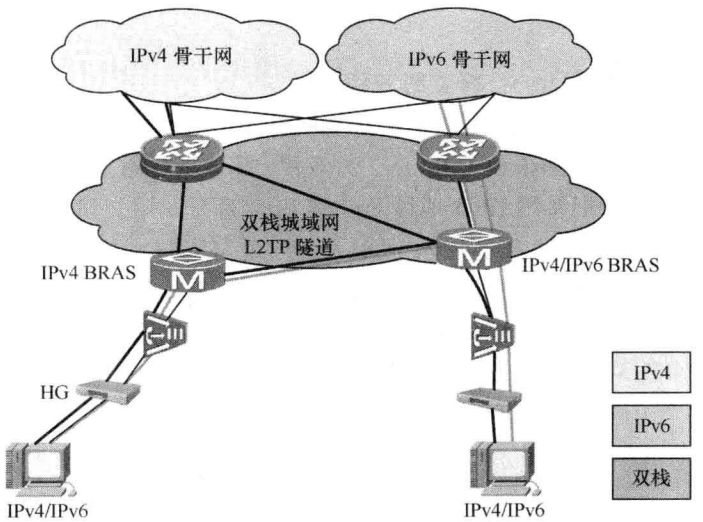


图 5-8 L2TP 隧道接入双栈技术

此方案的应用场景如下。

① 适用于 IPv4 BRAS 下挂 IPv6 用户，IPv6 用户访问目的地为 IPv6 骨干网。

② 城域网中有部分 BRAS 已经完成双栈改造。

③ L2TP 可以完全替代 6RD。

④ 现网 IPv4 BRAS 建立 L2TP 隧道到双栈 BRAS，将双栈用户发送到双栈 BRAS 认证，双栈 BAS 给终端分配 IPv4/IPv6 双栈地址。

⑤ 终端及业务为双栈或 IPv4 单栈。

在 IPv4 向 IPv6 过渡的初期，可以利用 L2TP 技术在 LNS 侧分配地址的特点，新增支持 IPv6 的 BRAS，将 PPP 的 IPv6 或者双栈用户转发到新增的 BRAS 上处理。这种情况下，无需升级现网作为 LAC 的 BRAS 设备。

L2TP 隧道只能用于 PPPoE 接入场景。关于 L2TP 的技术方案，将在 5.2.4 小节中详细描述。

4. 6RD（IPv6 Rapid Development）：IPv6 快速部署

6RD 是 6to4 隧道技术的改进版本，可以使用运营商 IPv6 地址前缀（而不同于 6to4 局限于特定前缀（2002：：/16）），更方便商用部署。

6RD 方案需要在城域网中新增 6RD 网关，并升级改造家庭网关，以支持 6RD。现有城

域网中绝大部分 CR、SR 及 BRAS 等设备仍然可以保持 IPv4 单栈运行不变，无需升级改造以支持 IPv6。6RD 方案部署后，IPv4 业务访问不变，IPv6 业务访问通过在支持 6RD 的双栈家庭网关与城域网新增的 6RD 网关之间建立 IPv6 over IPv4 隧道承载，实现 IPv6 终端与 IPv6 业务跨越 IPv4 网络的互通，如图 5-9 所示。

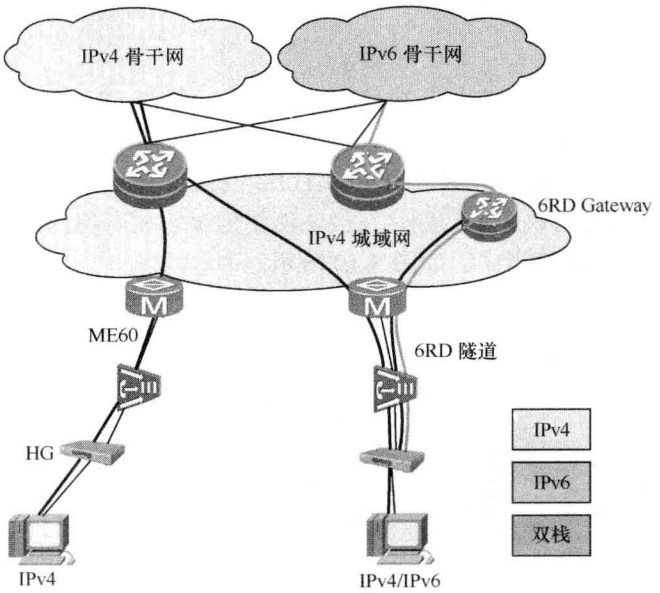


图 5-9　6RD 方案

6RD 方案的优势在于：可以实现在基本保持现有 IPv4 城域网络不变的前提下快速引入 IPv6，解决 IPv6 业务访问问题。但 6RD 方案中需要升级改造的 CPE 数量巨大，而且 6RD 方案对城域网并没有进行 IPv6 升级改造，后续向 IPv6 演进仍然需要完成大量投入，6RD 方案的初始改造成本和增量成本是必须考虑的因素。同时，由于 6RD 方案采用隧道技术，对于如何与 NAT 技术结合以缓解 IPv4 地址不足的问题，也没有成熟的解决方案。

综上所述，6RD 方案适用于 IPv4 公有地址充足的运营商，短期内没有 IPv4 地址不足的压力，希望通过现有 IPv4 网络为用户快速提供 IPv6 接入服务，在 IPv6 网络演进初期培育用户 IPv6 业务使用习惯的应用场景。但中国运营商的 IPv4 地址今明两年将消耗殆尽，面临业务发展的紧迫压力，部署 6RD 方案的最佳时间点已经过去，现阶段再选择 6RD 方案并不能解决实际问题。

此方案的特点如下。

① 6to4 技术的一种，能在现有 IPv4 网络基础上快速提供 IPv6。

② 适合的城域网为 IPv4，通常应用于 IPv6 孤岛间互连。利用 6to4，IPv6 站点可以不依赖于运营商 IPv6 基础而互通。

③ IPv4/IPv6 用户通过与 6RD Gateway 间的隧道访问 IPv6 Internet。

④ 终端及业务为双栈和 IPv4 单栈，家庭网关需要改造以支持 6RD。

⑤ 对家庭网关的二次改造是 6RD 最大的弊端。

6RD 方案不需要对现网的 BRAS 设备进行升级改造来支持 IPv6，但需要部署 6RD 网关，

当 IPv4 公有地址不足时仍然需要部署 NAT44 设备，而且用户 CPE 必须是双栈路由型，支持 6RD 能够发起建立隧道，适合还有较多 IPv4 公有地址储备的运营商，在过渡初期引入 IPv6 服务。有些运营商或迫于竞争、或因为客户要求，需要为用户提供 IPv6 服务，又不希望对现有网络进行较大改造，采用 6RD 是一个较为经济、风险较小的选择。6RD 对部分拥有较多 IPv4 公有地址储备的欧美运营商可能更具有吸引力，他们暂时还不希望对现有网络进行较大改动。

5. NAT64

NAT64 是一种有状态的网络地址与协议转换技术，一般与 DNS64 配合，实现 IPv6 主机到 IPv4 业务的访问。NAT64 方案适用于 IPv6 网络演进后期。随着 IPv6 单栈终端的大量出现，运营商通过在网络中部署 NAT64 设备为 IPv6 单栈用户提供访问未升级改造的 IPv4 单栈业务的服务。NAT64 涉及两种协议地址的翻译转换，网络运维需要考虑的问题比 NAT44 将更加复杂。相关标准目前还在讨论当中。如图 5-10 所示。

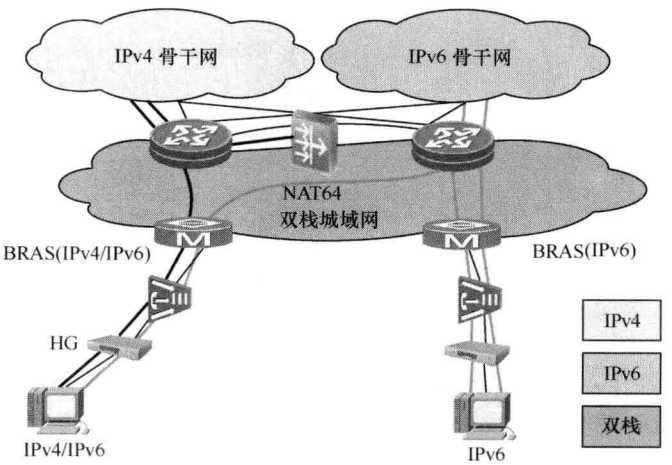


图 5-10　NAT64 方案

此方案的特点如下。

① 城域网已经完成双栈改造。

② IPv6 用户访问 IPv4 骨干网通过 IPv6-IPv4 网关完成。

③ 终端及业务：双栈和 IPv6 单栈。

④ 需要实现 DNS-ALG 用于 DNS 报文中的地址翻译。

⑤ 唯一已经产品化的翻译技术。

⑥ 只允许 IPv6 主动发起的连接。

5.2.4　接入网

在宽带接入网络中，不同运营商的业务模式不同，家庭网关运行的模式也不同。本节介绍 IPv6 接入方案的几个典型的应用场景。在这些接入应用场景中，宽带接入网络需要从 IPv4 升级到 IPv6。针对每一种接入方式，认证、授权和计费服务器都需要支持对应的 IPv6 属性。

家庭网关运行在终端模式时，可以为没有 IP 地址的终端发起连接，例如 VoIP 电话终端；也可以发起连接到网管，运营商通过该连接管理家庭网关设备。家庭网关运行在终端模式时，

其接入方式与其他终端设备（PC、机顶盒等）相同，此处不再详细说明。

1．家庭网关运行在桥接模式

图 5-11 所示为家庭网关（CPE）运行在桥接模式的用户接入场景，家庭中每个终端都需要单独认证并从运营商处获取地址。桥接模式接入时，CPE 只透传报文，无需升级支持 IPv6。

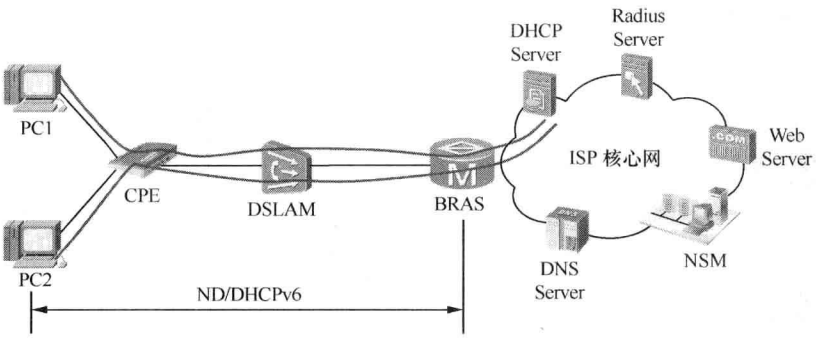


图 5-11　家庭网关运行在桥接模式

用户采用 PPPoX 的方式接入，对现网的改造要求最少，接入网都无需升级改造，只需要终端和 BRAS 设备支持 IPv6 即可。采用 L2TP 隧道做业务批发时，可仍使用现有的 IPv4 网络的 L2TP 隧道来支持 IPv6 业务批发，只需要作为 LAC 和 LTS 设备的 BRAS 支持 IPv6 报文转发，LNS 设备支持 IPv6 用户接入即可，如图 5-12 所示。

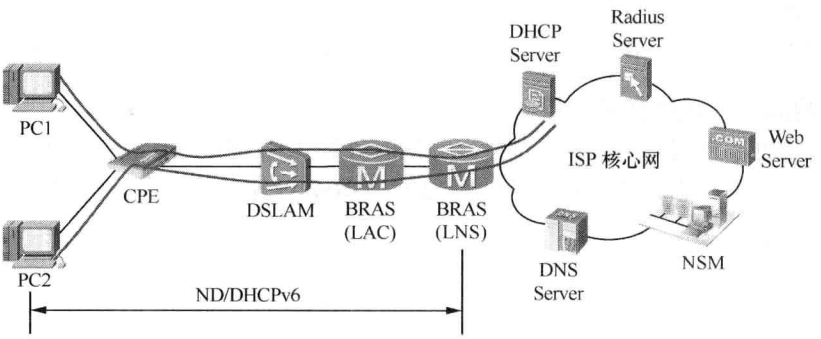


图 5-12　用户采用 PPPoX 方式接入时支持采用 L2TP 隧道接入

用户采用 IPoX 的方式接入，接入网设备必须支持 IPv6 协议多播报文的转发。采用绑定方式认证时，还需要支持 DHCPv6 LDRA 功能。

Windows 操作系统从 Windows Vista 开始，支持 PPPoE IPv6 拨号功能，目前已经部署 IPv6 的长沙电信和无锡电信，都采用了 PPP 认证，桥模式接入的方式支持双栈用户接入。

Windows XP 操作系统不支持 PPPoE IPv6 拨号和 DHCPv6 协议，只能通过非 PPP 封装的 ND 协议获取 IPv6 地址，适合采用 IPoX 802.1x 认证方式接入 IPv6 网络。

绑定认证适用于 IPTV 场景，不需单独输入用户名和密码进行认证，可以将多播复制点放到最接近用户的接入网设备上，降低了接入网中带宽的消耗。

Web 认证适合家庭上网用户。IPv6 单栈接入时，Web Server 需要升级支持 IPv6；双栈用

户接入时，则可以不升级 Web Server，用户仍然使用 IPv4 地址作 Web 认证。认证通过后，IPv4 和 IPv6 地址都拥有了访问网络的权限。

2. 家庭网关运行在路由模式

家庭网关运行在路由模式时，由家庭网关发起请求，运营商对家庭网关进行认证，为家庭网关分配 IPv6 前缀，家庭网关使用该前缀为家庭中有需要的终端分配 IPv6 地址。这种情况下，运营商只管理家庭网关。家庭内部的终端由家庭网关管理，场景如图 5-13 所示。

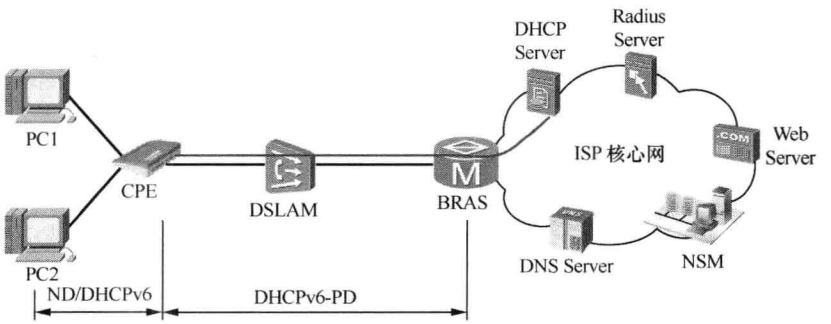


图 5-13　家庭网关运行在路由模式

采用 CPE 路由模式接入时，必须升级和替换用户家中的家庭网关设备，以支持 IPv6 接入。同桥接模式一样，用户若采用 PPPoX 的方式接入，则不需要升级接入网；若用户采用 IPoX 的方式接入，则接入网需要升级，以支持 IPv6。

家庭网关采用 WAN 口由 BRAS 分配地址的 numbered 路由模式接入时，建议使用 DHCPv6 协议为 WAN 口分配地址，在一个 Session 中同时完成 IPv6 地址和 IPv6 前缀的分配。

3. 家庭网关同时运行在多种模式

家庭网关也可以同时运行在多种模式，例如，PC 上网业务运行在路由模式，IPTV 运行在桥接模式，家庭网关自己作为终端运行 VoIP、TR069（家庭网关的网管业务）业务等，其应用场景如图 5-14 所示。

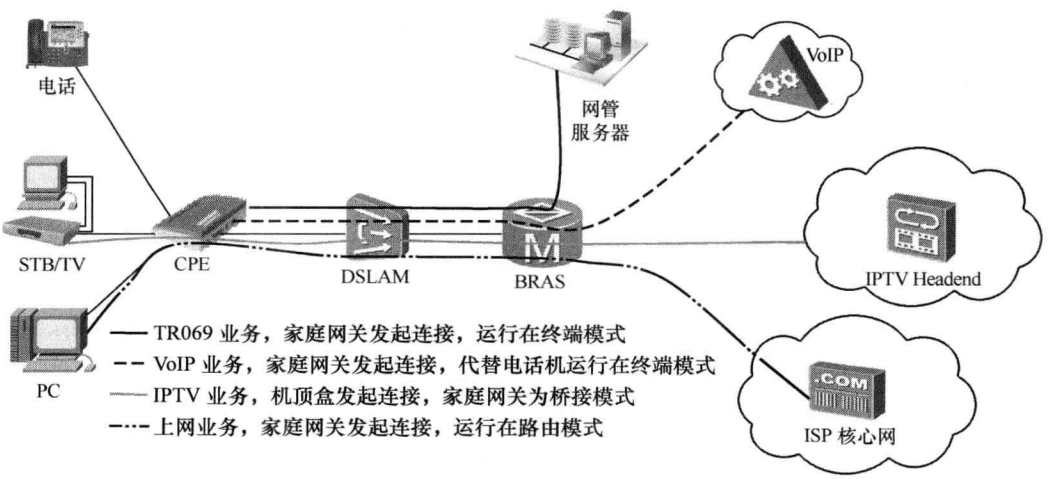


图 5-14　家庭网关运行在多种模式

5.2.5　IPv6 网络规划

运营商有多种业务，每种业务都有自己的特点，在 IPv4 向 IPv6 升级的过程中，需要综合考虑各种业务的特性，采用适合的升级方案。

1．按业务场景采取不同的 IPv6 过渡技术

网管、IPTV 等业务提供商较少，且用户间无互动的业务，可以直接升级到 IPv6；VoIP 等用户间互动模式单一，且业务提供商提供的业务较少，也可以直接升级到 IPv6；家庭宽带上网业务提供商众多，用户间互动复杂的业务则必须考虑配合不同的过渡技术（双栈，隧道和翻译技术，具体应用参考 IPv6 过渡技术和中国区部署建议），以支持用户同时访问 IPv4 和 IPv6 网络，如图 5-15 所示。

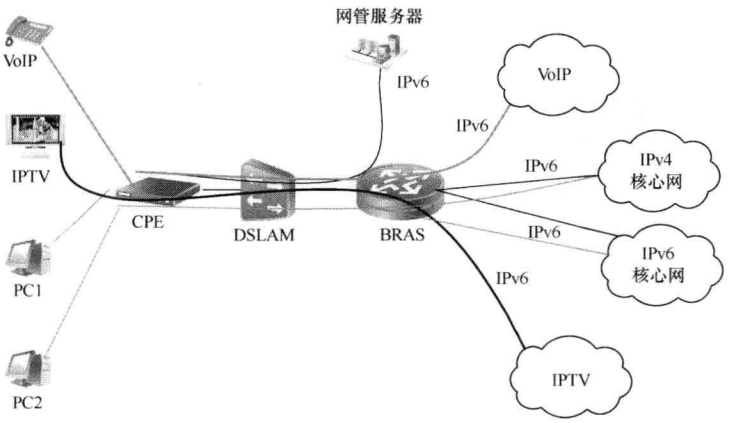


图 5-15　不同业务采用不同的升级方案

由上面的分析可以看出，除了宽带上网业务外，其他的业务过渡技术都相对比较单一。在宽带接入网络中，IPv4 升级到 IPv6，只是用户的地址从 IPv4 升级到 IPv6，用户的认证、授权和计费机制并没有变化，已有业务模型也没有变化。在 IPv4 向 IPv6 升级的过程中，需要考虑到不同业务的特性。例如，IPTV 和 VoIP 可以直接升级到 IPv6 网络，而用户上网业务需要和不同的过渡技术配合，使得用户可以同时访问 IPv4 和 IPv6 网络中的业务。

2．结合中国区现状，宽带上网业务演进方案

基于各种演进方案的分析，结合中国运营商的网络现状，适合中国运营商的演进路线如下。

首先启动现有网络的双栈平滑升级。对于无法升级支持 IPv6 的 BRAS 下的用户采用 L2TP 隧道接入双栈网络，采用双栈方式发展用户，培养用户 IPv6 使用习惯，积累运营商网络运营经验，推进 IPv6 业务发展。

在双栈部署的基础上，在 DS-Lite 成熟前，为 IP 地址紧缺地区存量设备部署 NAT444，节省 IPv4 公网地址。

DS-Lite 技术成熟后，在 IP 地址紧缺地区规模部署 DS-Lite 演进方案，将部分可升级 NAT44 设备升级为 AFTR 设备，减少设备投资。

当 IPv6 业务丰富后，发展 IPv6 单栈用户，并在网络中部署 NAT64 设备，实现 IPv6 单栈用户对少数未改造的 IPv4 单栈业务的访问，最终建成纯 IPv6 的下一代互联网，如图 5-16 所示。

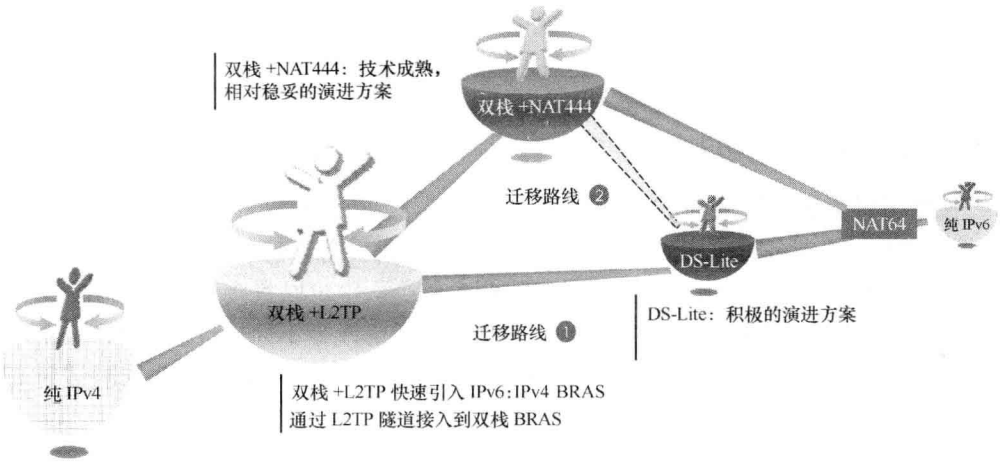


图 5-16　宽带上网业务演进方案

3．网络演进部署建议

（1）阶段一：早期部署阶段

IPv6 部署应遵循“先核心、后边缘”的部署原则。首先骨干网/IGW 支持双栈，根据现有网络部署技术和运维需求可选择 Native 双栈和 MPLS 6PE/6vPE 两种路线，如果现有网络已经支持 MPLS，选择 MPLS 6PE/6vPE 则无需对现有骨干 P 路由器做更改，如图 5-17 所示。

早期部署的业务需求主要来自于 VPN 专线。通过升级或新建 PE 路由器，可满足要求支持 IPv6 专线客户的需求。

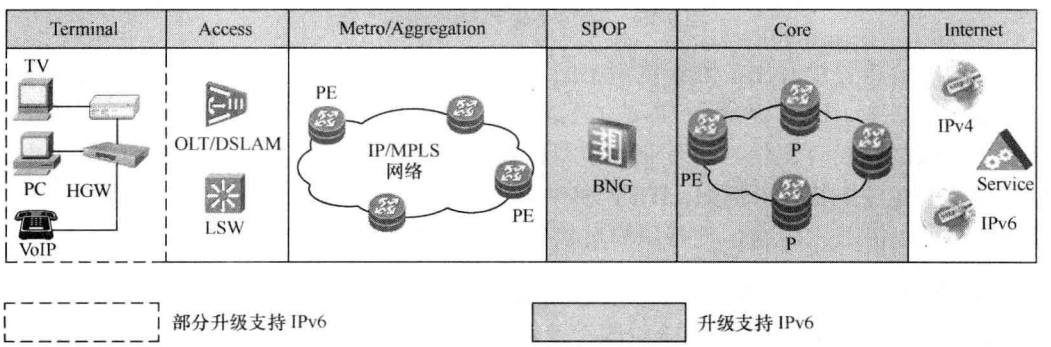


图 5-17　早期部署阶段，升级骨干网，改造 S-POP 和 CPE

SPOP 是 IPv6 公众宽带业务部署的核心和难点。由于 PPPoE 业务无需改造接入网，HSI 早期部署建议首先从 PPPoE 宽带用户开始。通过升级或新建 BRAS 以及认证计费系统，支持个人用户双栈接入。对于桥接式用户，不需要更换 HGW；而对于路由型用户，则需要更新为支持双栈的 HGW，更换的 HGW 应可以通过软件升级，支持后续演进。

对 VAS（Value Added Service）系统的升级改造应同时进行，以增加 IPv6 业务流量。

（2）阶段二：应对地址短缺

目前 IPv4 业务仍占据主流，而 IPv4 地址短缺成为多数运营商需要面对的问题，NAT 及其变种是解决地址短缺问题的必然选择。一般地，引入 NAT 会导致网络流量模型变化，使得

流量管理、用户业务识别的难度增加，甚至使得某些业务受损。

BNG/CGN 融合方案，通过 BNG 集成 DS-Lite 或 L2 Aware NAT，业务只需一次 NAT。分布式 NAT 结构降低了性能需求，容易隔离故障和定位。运营商无需管理终端的 IPv4 地址，无需部署 LOG，从而降低了部署成本，提高了用户感知。缺点在于：如果不开展 IPv6 业务，更换终端的成本会比较高；该方案依赖 BRAS 厂商的支持；基于 IP 地址识别用户的流量管理、DPI、VAS 将不能正常工作，如图 5-18 所示。

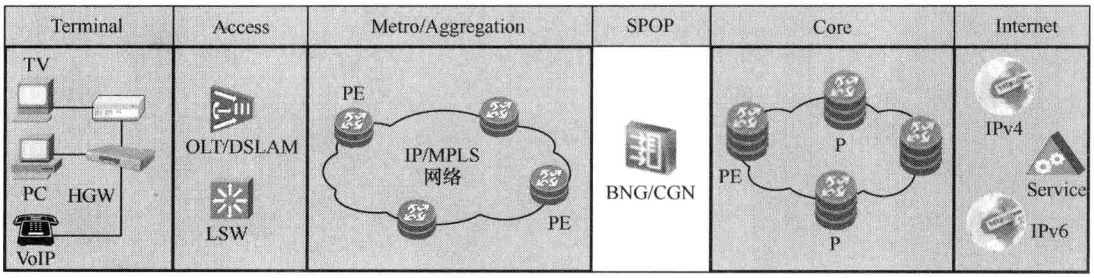


图 5-18　部署 BNG/CGN 应对地址短缺

集中式 NAT 目前已经被广泛部署。在网络出口放置集中式 NAT，对私网地址用户的业务进行重定向，到 NAT 设备上进行地址转换。独立式 NAT 部署的优点在于：设备可获得性好，原有网络和业务系统几乎不需要做更改。缺点在于：集中式设备故障影响面广，且故障定位较困难；需要建设溯源系统；对 IPv6 部署没有明显促进。

在双栈网络中，解决地址短缺依然可以采用集中式 NAT 方式，相当多运营商选择双栈 + NAT 作为演进部署方案，如图 5-19 所示。

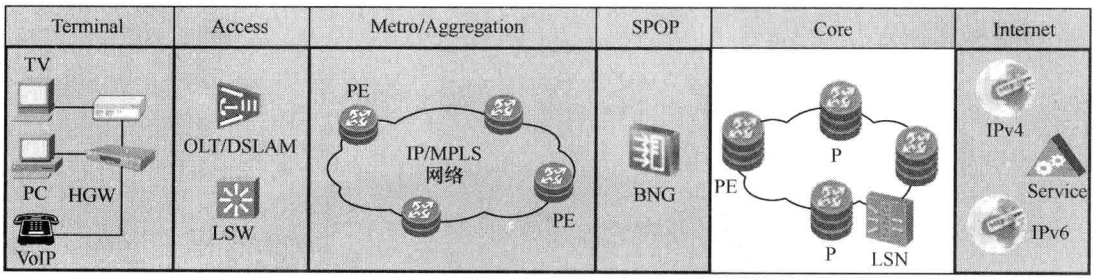


图 5-19　部署集中式 NAT 应对地址短缺

业界已经达成共识，认为 DS-Lite 是同时解决地址与加速 IPv6 发展的最佳方案，其相关标准已经成熟，被多个顶级运营商选中作为演进路线，成为 Native 双栈方案之外唯一的主流演进技术，终端和设备供应商已有商用产品发布。

两种主流的演进路线是可以兼容的。在初期部署阶段可采用 Native 双栈，同时进行 DS-Lite 试验（DS-Lite 只使用网络的 IPv6 协议栈），在条件具备的情况下可逐步切换到 DS-Lite 方案，然后逐步关闭承载网络的 IPv4 协议栈（保留 CGN 和 IGW），使得整体网络的投资成本最优。

（3）阶段三：向纯 IPv6 网络过渡

随着用户管理逐步向 IPoE 迁移，接入网支持 IPv6 业务的感知需求逐渐迫切。在 IPoE 接入场景下，IPv6 协议报文一般为特定多播报文，而且原有 DHCP option、DHCP Snooping、

IP/MAC 绑定、多播 Snooping/proxy 等安全特性都需要针对 IPv6 进行开发。此外，在 IPv6 协议下，ICMPv6 成为接入层的新攻击点，Neighbor Discovery Snooping 成为接入层设备必备的安全特性。

随着 IPv6 网络部署的推进，对成本、功耗、海量、部署简单的需求将导致纯 IPv6 终端爆炸性增长，但部分终端可能仍存在访问 IPv4 业务的需求。通过部署支持 NAT64 的电信级 NAT，可使得 IPv6 终端以可管理的方式访问 IPv4 网络和业务，如图 5-20 所示。

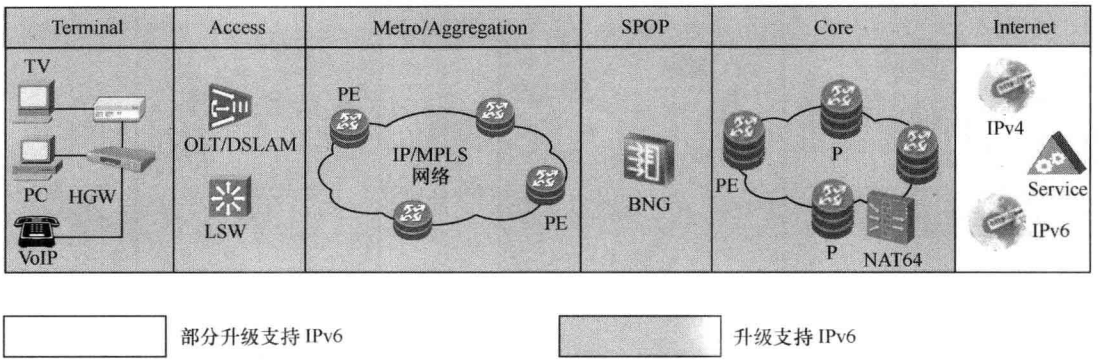


图 5-20　升级接入层和 CGN 部署纯 IPv6

5.3　业务部署

5.3.1　业务地址需求分析

（1）当前中国电信 IPv4 地址使用情况

如图 5-21 所示。

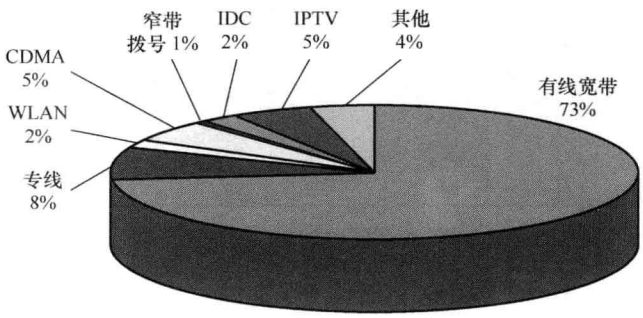


图 5-21　中国电信 IPv4 地址使用情况

① 有线宽带等固网业务占用的 IPv4 公网地址最多，占用 7450 万左右。

② 专线接入等大客户业务占用 IPv4 公网地址 810 万左右。

③ 移动宽带 CTWAP 拨号采用的是私网地址；CTNET 拨号采用的是公网地址，占用 550 万左右。

④ IPTV 业务存在使用公网地址和私网地址两种模式，其中公网地址占用 470 万左右。

（2）各个业务对 IPv4 公网地址消耗情况

见表 5-3。

表 5-3　　IPv4 公网地址消耗情况

业务	当前用户数量	当前公/私网	公网地址需求量	当前建议
固定宽带（拨号用户）	海量	公网	需求量大	双栈 + NAT
固定宽带（专线用户）	较少	公网	较少	IPv4/IPv6
IPTV（服务器）	较少	公网	较少	IPv4/NAT64
IPTV（机顶盒）	较多	封闭系统，公网复用/私网地址	较少	IPv4/IPv6
移动宽带	海量	私网	较少	IPv4 私网/ IPv6
IDC	较少	公网	较少	IPv4/NAT64
物联网	无	无	海量	IPv6

由表 5-3 可知，当前对于 IPv4 公网消耗最大的业务就是固定宽带拨号用户。所以对于网络改造，当前最紧急的也是固定宽带拨号用户。

5.3.2　宽带接入业务

宽带上网业务是对公网消耗最大的业务，也是当前最需要改造的业务。改造建议如下。

① 对于固网宽带和 IPTV 业务可逐步以省为单位引入私网地址，为新增用户分配私网地址，老用户暂时不迁移。

② 移动宽带业务由于存在漫游特性，保持现有模式，可全国统一考虑私网地址的部署。

③ 原有软交换、IMS 网络保持 IPv4 私有地址不变，新增的 SIP 话机等可与固网宽带采用同样的私网地址，通过地址转换方式接入 BAC。

下一代互联网 IPv6 的发展不单单是网络层面的演进，IP/IT 支撑系统需要同步支持各项 IPv6 业务功能。固网宽带用户业务受理流程牵涉很多环节，如 CRM 系统、SPS 系统、TSAP 系统、计费系统、AAA 系统等，如图 5-22 所示。

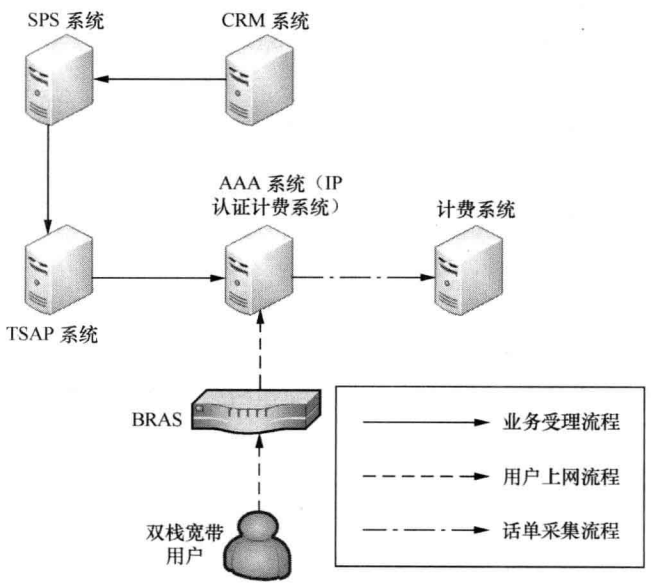


图 5-22　宽带业务受理流程图

CRM 系统（进行业务受理）→SPS 系统（进行业务开通）→TSAP 系统（进行局端设备数据自动数据激活）→AAA 系统（认证及计费）→BSN 系统（完成 IPv6 计费 CDRs 的批价与账务）。

现有网络中的 IP/IT 支撑系统已具备支持 IPv6 宽带业务接入的能力，为更好推动 IPv6 的演进，未来的发展需切实关注和解决的问题如下。

① 由以往的单栈 IPv4 宽带业务转变为单栈 IPv4、双栈、单栈 IPv6 等多元化宽带业务，支撑系统针对不同业务属性用户的受理与切换等问题。

② 过渡技术的引入，要求支撑系统做好功能评估，为后续工作提前筹备。

③ 运营商都拥有各自的自营业务，过渡技术的应用所带来的业务影响，支撑系统应积极应对，寻求合理有效的解决方法。

5.3.3 无线宽带（含 3G 和 WLAN）业务

无线宽带业务在这里指用户可通过 PC 终端上的 WiFi 网卡或 3G 上网卡，选择无线宽带（WLAN）、无线宽带（3G）等不同网络接入方式上网的业务。手机终端的上网业务不在本章节范围内做介绍。

1. WLAN 接入产品

目前，用户上网通过 WLAN 接入到城域网 BRAS 设备，由 BRAS 设备为用户分配公网地址，用户发起 HTTP 访问请求后被重定向到 WLAN Portal 进行认证。用户输入用户名和密码后，Portal 将用户认证信息转发到 AAA 进行认证后，用户可以通过认证后的公网地址访问互联网。

事实上，WLAN 接入方式下的无线宽带业务在网络侧与固网宽带业务方式基本一致。因此，除了针对 WLAN Portal 的额外改造需求外，对 IPv6 的支持同样要求城域网网络设备支持 IPv6 接入，IP、IT 支撑系统支持 IPv6 用户属性的识别和处理。

WLAN 的改造过渡可分为如下两个阶段。

① 一阶段，支持用户使用 IPv4 HTTP 进行 Portal 认证，认证通过后可以使用双栈访问 Internet。支持 IPv6 话单信息生成。

② 二阶段，在 BRAS 设备支持 IPv6 HTTP 重定向情况下，支持用户通过 IPv6 方式访问 Portal。

无线宽带在 WLAN 方式下，地址分配是 BRAS 设备实现的。因此，应从地址枯竭情况出发，通过合理规划 BRAS 的地址配置来控制是否引入地址复用技术。

2. 3G 接入产品

目前业务形态为用户使用插有上网卡的 PC，并在 PC 上安装、运行拨号客户端软件，选择无线宽带（3G）或无线宽带（1x）接入方式拨号上网。

上述业务要实现对 IPv6 接入的支持，对终端和网络有以下要求。

① 终端侧，用户的 PC 地址如需获取 IPv6 地址，并通过 IPv6 访问 Internet，应使用 Windows Vista 以上版本的操作系统，同时需要在操作系统内开启 IPv6 协议栈。

② 网络侧，由于业务与手机终端上网一致，都是通过移动分组网进行承载的，需要 PDSN、移动 AAA 等网络和支撑系统支持 IPv6 用户的业务开通、使用、计费等流程。网络侧的过渡部署可参见“5.3.4 其他业务”的“移动承载网”部分。

综上，无线宽带业务的过渡演进应以双栈技术为主，通过网络、支撑系统的升级改造支撑 IPv6 业务应用。在此基础上，结合地址枯竭情况合理引入过渡技术，缓解地址压力。

5.3.4　其他业务

1. IPTV

借助国家三网融合政策的推动，截止到 2010 年年底我国 IPTV 用户已经发展到 800 万，未来几年将持续大规模发展，IPTV 用户量将以倍数关系增长，由此将产生海量 IP 地址需求。

考虑到 IPv4 公网地址已经消耗殆尽，未来网络必将向 IPv6 演进的趋势，以及 IPTV 业务系统的相对封闭性，使得进行 IPv6 升级改造相对简单。为了降低运营商的多次投资，减少由 IPv4 向 IPv6 转换所带来的复杂工作量，使得 IPv6 IPTV 业务的试验与规模发展成为新的亮点，不仅解决了对 IP 地址的海量需求，而且推动了 IPv6 端到端产业链的成熟，如图 5-23 所示。

IPv6 IPTV 端到端部署需要综合考虑多方面因素，IPTV 平台、骨干网、城域网 CR、BRAS、SR、接入网及机顶盒都需支持双栈。初期部署可考虑使用单播承载，后续随着客户数量的增长、IPTV 流量的增大，采用多播承载，引入 IPv6 多播复制点下移技术，降低城域网压力。

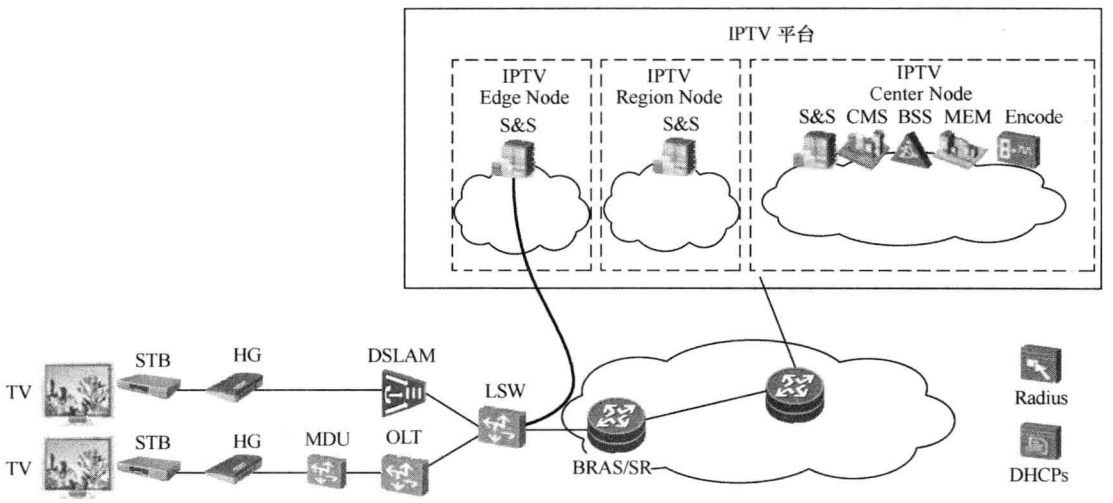


图 5-23　IPTV 网络拓扑结构图

2. IDC 网

IDC 是互联网内容的聚集地，而公网 IP 地址是提供 IDC 业务的基本条件，IPv4 地址的枯竭对于 IDC 业务正常运营有着最直接的影响。

经过多方试验和研究，IDC 网已具备升级为双栈网络的条件，可以为接入 IPv6 的 ICP 服务商提供很好的平台。无可否认，网络现状仍然以 IPv4 互联网业务为主，很多 ICP 服务提供商升级改造为 IPv6 或双栈环境的积极性不高。一方面升级改造带来一定风险，ICP 服务提供商担心会影响到现网业务；另一方面没有 IPv6 宽带用户客户源，无法体现 IPv6 网络的优势所在；其次，ICP 服务提供商不想投入太多资金在 IPv6 上，因为在短期内无法看到收益。

基于以上多种原因，目前大量 ICP 仍然以 IPv4 接入为主。为了打破僵局，应以政府为主

导，联合产业链各方，尽快制定内容和应用向 IPv6 迁移的整体方案，解决应用层面 IPv4/IPv6 的互通问题，为用户提供丰富的 IPv6 内容源，调动用户向 IPv6 网络迁移的积极性。在 IDC 网内可考虑采用 IPv4-IPv6 的协议转换技术，实现 IPv6 网络用户也能正常访问 IPv4 网络中的业务内容，比如现今业界广为熟知的反向代理、NAT64、IVI、应用层代理等技术，从而进一步推动 ICP 服务提供商往 IPv6 方向迈进。

3. 移动承载网

目前分组域承载移动互联网业务，对地址需求量很大，也需要率先向 IPv6 过渡。而电路域和无线网络都属于封闭网络，对公有地址需求不大，暂时无需升级为 IPv6 网络，设备之间的管控可继续使用 IPv4 地址。

现有网络中移动网设备已具备双栈、单栈 IPv6 的接入能力。随着演进步伐的迈进，公网 IPv4 地址已不能满足业务需求。但由于 IPv6 的引入处于初级阶段，再加上无实际 IPv6 业务内容可供访问，所以 IPv4 仍占主导地位。预计移动承载网的未来发展将以两条主线为指引：一方面积极开展 IPv6 的深入研究，全面系统地进行功能与性能评估，实施穿越流程的业务试验，包括本地业务和漫游业务；另一方面引入过渡技术，缓解公网 IPv4 地址资源不足的问题。除此以外，市场终端方面也需要积极开拓手机终端等产品对 IPv6 支持的力度，为日后新型业务的面世做好合理规划。

网络层面，以 IPv4 为主导，IPv6 为目的，双栈为基础，通过各种过渡技术和转换技术实现 IPv4 到 IPv6 网络的平滑过渡；业务层面，做好充分准备，从 IP/IT 支撑系统、业务系统到终端，完善 IPv6 相关功能的升级改造，为后续工作的开展未雨绸缪。以现网业务影响小、成本投入低为前提，尽可能保持现有网络结构，实现端到端业务体系和网络架构向以 IPv6 为基础的下一代互联网平滑演进。

5.4 IPv6 网络地址规划

5.4.1 设计思路

一般情况下，IP 地址通过网络标识 + 主机标识来表示一个唯一的网络设备（接口），设备（接口）有了 IP 地址，才有可能充分利用 IP 协议实现网络互通以及应用层的处理。也可以说，IP 地址规划是实现网络功能的基础。

1. IPv4 地址

IP 地址分为 IPv4 地址和 IPv6 地址。

IPv4 地址长度为 32bit，由 Net-ID 和 Host-ID 组成，分为图 5-24 所示的 5 类。

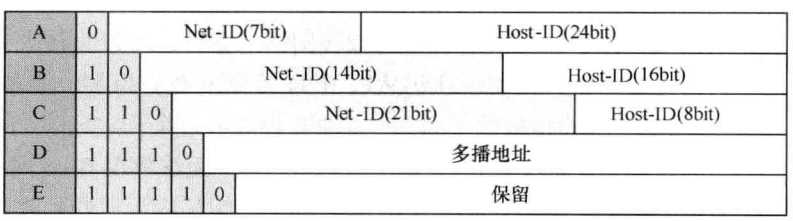


图 5-24 IPv4 地址

在实际使用过程中，有一些特殊的 IP 地址，其范围和描述见表 5-4。

表 5-4　　特殊 IP 地址

IP 地址编号	IP 地址子网号	IP 地址主机号	可否为源地址	可否为目的地址	描述
全“0”	—	全“0”	Yes	No	用于网络上的主机
全“0”	—	主机号	Yes	No	用于网络上的特定主机
127	—	任何值	Yes	Yes	用于环回地址
全“1”	—	全“1”	No	Yes	用于受限的广播（永远不被转发）
Net-ID	—	全“1”	No	Yes	用于向以 Net-ID 为目的的网络广播
Net-ID	SubNet-ID	全“1”	No	Yes	用于向以 Net-ID、SubNet-ID 为目的的子网广播
Net-ID	全“1”	全“1”	No	Yes	用于向以 Net-ID 为目的的所有子网广播

为了解决 IP 地址短缺的问题，提出了私有地址的概念。私有地址是指内部网络或主机地址，这些地址只能用于某个内部网络，不能用于公共网络。RFC 1918 描述了为私有网络预留的 3 个 IP 地址段，见表 5-5。

表 5-5　　私有 IPv6 地址

网络类型	地址范围	说明
A	10.0.0.0～10.255.255.255	一般用于 CPE NAT 地址池或大型企业内部
B	172.16.0.0～172.31.255.255	一般用于中小型企业内部
C	192.168.0.0～192.168.255.255	一般用于家庭网络或小型企业内部

私网地址不能作为 Internet 路由发布到公网上去，私网地址一般会形成 VPN 私网路由，在公网中通过 MP-BGP 协议在 VPN 隧道里传递。

2．IPv6 地址

IPv6 地址长度为 128bit，由前缀和接口标识组成。前缀相当于 IPv4 地址中的 Net-ID，接口标识相当于 IPv4 地址中的 Host-ID。前缀一般由运营商从地址管理机构申请后分配给终端用户，接口标识可以通过 MAC 地址转换、设备随机生成或手工配置生成。

IPv6 地址分为单播地址、多播地址和任播地址，分别用于不同场景。

（1）IPv6 单播地址

标识一个接口，目的为单播地址的报文会被送到被标识的接口。

根据 RFC 3513 的定义，规定对于单播地址有一个总的原则：若单播地址前 3 位不以 000 开头，接口 ID 必须为 64bit，而那些前 3 位以 000 开头的单播地址就没有这个限制，详见表 5-6。

表 5-6　　IPv6 单播地址

分类	地址范围	说明
未指定地址	全 0 地址::	接口未分配 IPv6 地址前的地址
环回地址	::1	相当于 IPv4 的 127.0.0.1 环回地址

续表

分类	地址范围	说明
全球单播地址	2001::/16	IPv6 Internet 地址，用得最多的地址
	2002::/16	6to4 隧道（IPv6 over IPv4）地址
	2003::/16～3ffd::/16	未分配
	3ffe::/16	6bone 地址
内嵌 IPv4 地址的 IPv6 地址	::10.153.70.200	IPv4 兼容 IPv6 地址，以 96 位 0 加上 32 位 IPv4 地址形成，用于 IPv4 兼容 IPv6 自动隧道
	::FFFF:10.1.2.6	IPv4 映射 IPv6 地址，以 80 位 0 加上 16 位 1，再加上 IPv4 地址形成，用于 IPv4 与 IPv6 的互通，目前有 SIIT（Stateless IP/ICMP Translation）应用，参考 RFC 2765
	64:FF9B::10.1.2.7	Stateless NAT64 应用地址，用于 IPv6 Host 访问 IPv4 Server，注意在 NAT64 应用下，SP 可以手动修改前缀
链路本地地址	以 FE80 开头	只在本地二层域内生效，FE80 加上接口 ID 自动生成，一般无需配置，用于完成本地设备间最基本的通信，如 ND
唯一本地地址	以 FEC0 开头	相当于 IPv4 的私网地址

（2）IPv6 多播地址

标识多个接口，多播地址的报文会被送到被标识的所有接口。

IPv6 的多播与 IPv4 相同，也就是发往多播目的地址的报文会被发送到该多播地址代表的一组接口。

在 IPv4 中，多播地址的最高三位设为 111。在 IPv6 网络中，多播地址也有特定的前缀来标识，其最高位前 8 位为 1。图 5-25 显示了多播地址的结构。

图 5-25　IPv6 多播地址格式

flags（标志）字段有 4bit，目前只使用了最后 1bit（前 3 位必须置 0）。当该位值为 0 时，表示当前的多播地址是由 IANA 所分配的一个永久分配地址；当该值为 1 时，表示当前的多播地址是一个临时多播地址（非永久分配地址）。

scope（范围）用来限制多播数据流在网络中发送的范围。该字段占有 4bit。RFC 3513 是这样定义该字段的。

0：预留；

1：节点本地范围，单个接口有效，仅用于 Loopback 通信；

2：链路本地范围；

4：管理本地范围，由管理员配置；

5：站点本地范围；

8：组织本地范围，属于组织的多个站点；

E：全球范围；

F：预留。

其他值暂时没有定义。

从这里我们就可以推断出，FF02::2 是一个链路本地范围的多播地址，而 FF05::2 是一个站点本地范围的多播地址。

group ID（组 ID）字段长度为 112bit，用以标识多播组。我们知道 112bit 最多可以生成 2112 个组 ID。目前，RFC 2373 并没有将所有的 112bit 都定义成组标识，而是建议仅使用该 112bit 的最低 32bit 组 ID，将剩余的 80bit 都置 0。如图 5-26 所示。

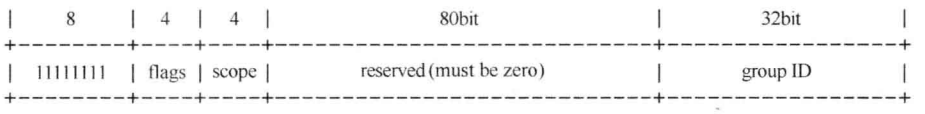


图 5-26　组 ID 使用方式

这样可以每个组 ID 都映射到一个唯一的以太网多播 MAC 地址（参见 RFC 2464）。

类似于 IPv4，IPv6 同样有一些特殊的多播地址，这些地址有特别的含义，这里举几个例子（还有很多类似的特殊地址）：

FF01::1（节点本地范围多播地址）；

FF02::1（链路本地范围所有节点多播地址）；

FF01::2（节点本地范围所有路由器多播地址）；

FF02::2（链路本地范围所有路由器多播地址）；

FF05::2（站点本地范围所有路由器多播地址）。

在 IPv6 多播地址中，有一种特别的多播地址，叫做 Solicited-node 地址，即被请求节点多播地址。被请求节点多播地址是一种特殊用途的地址，主要用于重复地址检测和节点获取相同本地链路上邻居节点的链路层地址。

被请求节点多播地址由前缀 FF02::1：FF00:0/104 和单播地址的最后 24bit 组成。一个单播地址对应一个被请求节点多播地址。被请求节点多播地址受限范围为本地链路范围。

在 RFC 3513 中定义了一些多播地址：

FF00::～FF0F::为保留地址，不允许被分配；

FF01::1、FF02::1 代表所有节点；

FF01::2、FF02::2、FF05::2 代表所有路由器；

FF02::1:FFXX:XXXX 被分配给被请求节点多播地址。

（3）IPv6 任播地址

标识多个接口，任播地址的报文会被送到最近的一个被标识接口，最近节点是由路由协议来定义的。

这是 IPv6 特有的地址类型，它用来标识一组网络接口（通常属于不同的节点）。目标地址是任播地址的数据分组将发送给其中路由上最近的一个网络接口。适合于“One-to-One-of-Many”（一对组中的一个）的通信场合。接收方只需要是一组接口中的一个即可，如移动用户上网就需要因地理位置的不同，而接入离用户最近的一个接收站，这样才可以使移动用户在地理位置上不受太多的限制。这和多播中的 Anycast RP 应用场景有些类似。

任播地址从单播地址空间中进行分配，使用单播地址的任何格式。因而，从语法上，任播地址与单播地址没有区别。被分配具有任播地址的节点必须得到明确的配置，从而确定它是一个任播地址。目前，任播地址仅被用做目标地址，且仅分配给路由器。

在 RFC 3513 中定义了子网路由器任播地址（Subnet-Router Anycast Address），其接口 ID 为全“0”。

发往该任播地址的报文会被发送到任播地址所代表的子网（子网路由器任播地址的前缀）内的某一台路由器，该路由器是离得最近的一台。所谓最近，一般是路由的概念。

5.4.2 设计原则

（1）连续性

IP 地址分配要尽量给每个地市城域网/区域分配连续的 IP 地址空间。在每个地市城域网中，相同的业务和功能尽量分配连续的 IP 地址空间，有利于路由聚合以及安全控制。通过 IP 地址规划尽量减少核心设备之间的穿越流量（流量大于 10GB 才考虑）。

（2）灵活性

IP 地址的分配需要有足够的灵活性，能够满足各种用户接入，如小区用户、专线用户等的需要。

（3）扩展性

考虑用户增长和业务增长，各管理域必须预留足够的空间，以防 5 年内再引入新的地址段。

（4）合理性

IPv4 地址的分配采用 VLSM（变长掩码）技术，以保证 IP 地址的利用效率。采用 CIDR 技术，这样可以减小路由器路由表的大小，加快路由器路由的收敛速度，也可以减小网络中广播的路由信息的大小。所有网络设备的链路接口和 loopback 接口都使用公网地址。IP 地址规划以区域分层为主，区域内按照业务模式划分地址块。封闭的承载网和封闭业务可以考虑使用 IPv4 私网地址或 IPv6 ULA 地址，或者使用独立的地址段。为了便于管理分配并易于聚合，IPv6 地址通常只分配几种固定长度的前缀；例如/64、/56、/48 等，避免分配各种长度的前缀。对于 IPv4 用户，如果公网地址不够，可以采用私网地址，在汇聚层提供网络地址转换 NAT 功能，支持公私网地址混合编址。若地址资源允许，也可以考虑给企业等大客户出口直接分配公网地址。

对于 IPv6 多播业务，建议一般采用 scope 为 8（组织本地范围，属于组织的多个站点）的多播 IPv6 地址。地址规划方案越早实施越好，避免导致后期产生过多的地址碎片。

5.4.3 设计要素

1. IP 地址模式

IPv6 没有私网地址的概念（但是 RFC 4193 又重新启用了 ULA 地址，和 IPv4 的私网地址是一个意思，一般用于 VPN Site 内部地址规划），IPv4 存在公网地址和私网地址两种类型。

（1）私网地址

建议采用 IETF RFC 1918 定义的私网地址网段（B 类地址），私网地址方式的业务采用

L3 VPN/VLAN 进行隔离。理论上每种业务可以独享整个私网地址空间，独立进行地址的分配，但不排除为管理上的原因对不同业务的地址段进行划分。

对于不需要访问外部资源的业务，可以采用私网地址方式。

对于需要访问公网资源的业务，如果地址资源需求量较大，也可以采用私网地址方式，这种情况下，访问公网资源则需要经过 NAT 设备进行公网和私网间的地址转换。

（2）公网地址

采用 SP 公网地址资源开展业务。由于公网地址资源较为紧张，需要控制采用公网地址方式的业务数量，大量消耗地址资源的业务尽量不要采用公网地址的方式给用户分配地址。

2. IP 地址分配方式

业务可以采用以下两种地址分配方式。

（1）静态配置方式

手工配置 IP 地址，不需要接入认证和地址分配过程，一般用于企业用户。对于手工静态配置 IP 地址的用户路由发布，有如下 3 种方式。

① IPv4 可以通过在 BRAS 上配置 Static-User 发布静态用户路由。

② 通过拨号触发 Radius 向 BRAS 发布 Framed-Route 用户路由。

③ 三层接入场景下，在接入设备配置静态路由即可。

注意：IPv6 静态 IP 用户接入情况下，必须要 CPE 支持 PD 才能够通过 BRAS 直接发布用户路由，否则需要 BRAS 支持 Framed-Route 通过 Radius 下发用户路由。

（2）动态分配方式

用户每次通过 IPoE 或 PPPoE 上线时，都会根据地址资源的状况，动态决定分配给用户的 IP 地址，可以根据用户特征灵活分配地址。

对于动态固定方式，需要保证用户每次上线均采用相同的 IP 地址；对于 IPoE 接入方式，可以在用户侧进行静态 IP 地址配置；对于 PPPoE 接入方式，则需要 BRAS、Radius 等配合，保证每次给用户分配已注册的 IP 地址。注意：动态固定分配的用户地址路由在 BRAS 上通过 Radius 下发 Framed-Route 发布，目前产品不支持在 BRAS 上直接配置 IPv6 静态路由，发布固定地址用户路由（IPv4 可以）。

地址分配给用户后，在用户上线期间，该地址会一直被其占用。当接入设备检测到用户下线后，动态申请到的地址会被释放回地址池，供其他用户使用。

3. IPv4 地址规划

（1）设备接口地址

设备接口地址用于网络互联，是建立转发路径的重要基础，一般采用静态配置，为公网地址。

（2）设备环回地址

设备环回地址用来标识某个设备的逻辑 ID，通常用于上层协议协商过程中的身份标识，采用静态配置，为公网地址形式。

（3）服务器地址

服务器是公共资源，需要提供给不同的用户访问，AAA Server、DNS Server 等服务器都需要采用静态方式分配公网 IP 地址。

（4）管理 IP 地址规划

管理 IP 地址均采用公网地址。

网络设备管理地址（用于 NMS 系统）采用静态分配的公网地址。

管理 VRRP 采用静态分配的私网地址。

RGW 管理采用动态分配的私网地址。

（5）HSI 业务地址

HSI 业务用户量巨大，一般采用动态 IP 地址分配，可能使用私网 IP 地址，也可能使用公网 IP 地址，在使用私网 IP 地址的情况下，需要通过 NAT 设备将私网 IP 地址转换为公网 IP 地址。如图 5-27 所示。

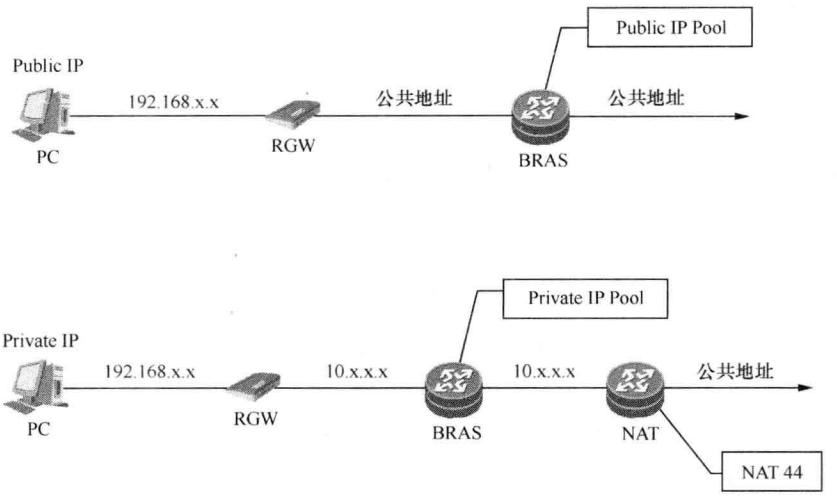


图 5-27　HSI 业务 IPv4 地址规划

（6）VoIP 业务地址

VoIP 为封闭式业务系统，采用 DHCP 方式获取私网 IP 地址。RGW 作为 DHCP Client 发起 DHCP 认证请求，DHCP Server 完成认证并分配私网 IP 地址，如图 5-28 所示。

图 5-28　VoIP 业务 IPv4 地址分配流程图

（7）IPTV 业务地址

IPTV 业务（通用组网场景）也为封闭系统。单播业务（如 VoD）一般采用 DHCP 方式获取私网 IP 地址；多播业务（如 BTV）采用 PIM-SM 或者 PIM-SSM 三层多播方式，多播组采用多播地址段（如 239.0.0.0/8，采用哪个多播地址段根据实际规划而定），如图 5-29 所示。

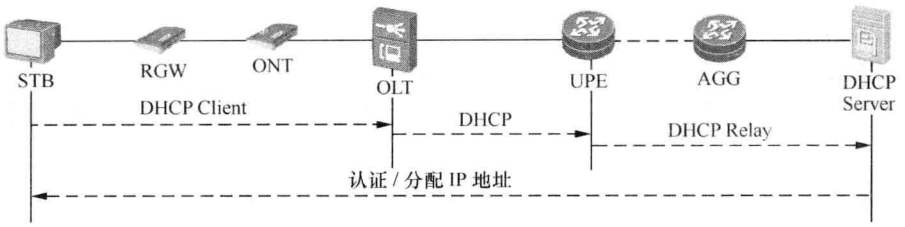


图 5-29　IPTV 业务 IPv4 地址分配流程

注：国内 IPTV 场景通过 SR 进行 DHCP Relay 申请 IP 地址。

（8）企业业务地址

中小企业（SME）Private IP 互联网专线业务采用私网地址，动态分配，IP 分配过程和个人 HSI 业务相同。

SME Public IP 互联网专线业务需要给企业分配指定数量的公网 IP 地址网段。

对于采用家庭网类（RGW）的接入方式，还需要为 RGW 的 WAN 口分配静态的私网地址。RGW WAN 口配置私网主要是为了节约 IP 地址，如图 5-30 所示。

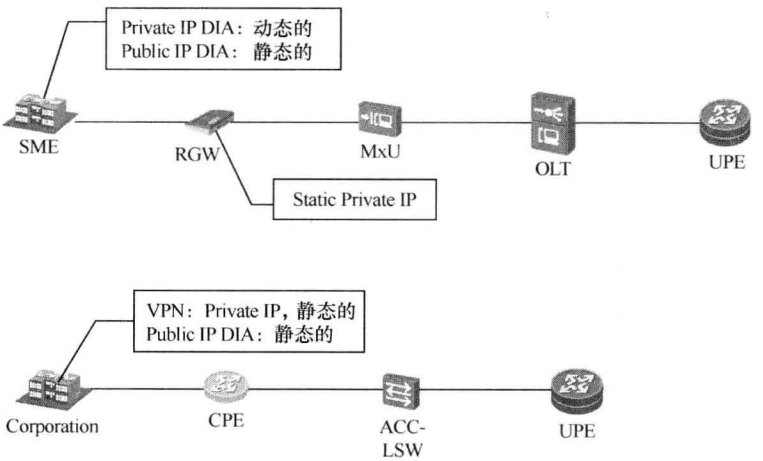


图 5-30　企业业务 IPv4 地址规划

大企业（Corporation）三层 VPN 业务采用私网地址，静态分配，L2 VPN 业务，包括 E-LINE/E-LAN 等 SP 网络只对报文进行二层处理，不涉及 IP 地址的分配问题。

Corporation Public IP 互联网专线业务采用公网地址，静态分配。

4．IPv6 地址规划

（1）地址分配方式

IPv6 网络中，设备接口地址、设备环回地址、服务器地址和 IPv4 一样，采用静态公网地址。对于管理 IP 地址和业务 IP 地址，IPv6 和 IPv4 区别不大，但在分配细节上有一些细微的区别。

① 管理 IP 地址都采用静态分配。

② 业务地址可以动态分配，也可以静态配置。

③ 特殊情况（如封闭系统）可以采用 ULA 地址。

（2）IPv6 前缀规划

IPv6 地址虽然很多，但也需要合理规划网段（前缀）以避免资源浪费，并简化管理。不同的用户对地址的需求数量是不一样的。在进行前缀规划的时候，尽量采用固定长度的前缀。

表 5-7 显示了各类用户的地址规划及子网前缀长度建议。

表 5-7　　IPv6 前缀规划

用户类型	用户网络属性	子网前缀	子网数量
家庭用户（桥接模式）	只有一个子网	/64	1
家庭用户（路由模式）	有多个子网	/56 或/60	256/16
中小型政企用户	有多个子网	/56	256
大型政企用户	有多个子网	/48	65535
设备互联	只有一个子网	/64	1

（3）IPv6 地址层次化

运营商在进行地址规划的时候，需要考虑地址的层次化结构。不同的网络位置，规划不同的前缀长度，如图 5-31 所示。

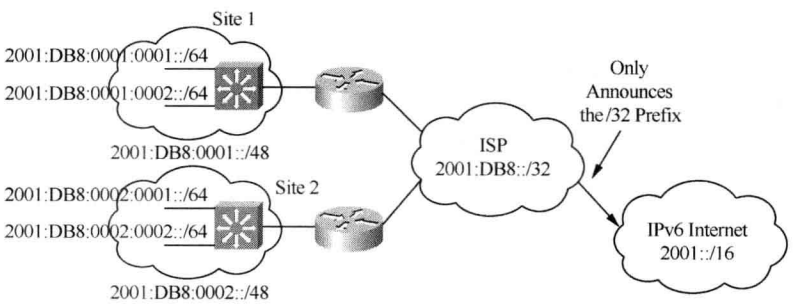


图 5-31　IPv6 地址层次化结构

这样做的好处主要有以下几点。

① 利于路由快速查找。

② 借助路由聚合，有效缩短路由表长度。

③ 提高路由器报文转发效率。

④ 便于网络隔离和路由受限发布，提高管理性和安全性。

（4）IPv6 多播地址规划

如果 IPTV 系统升级到 IPv6，对于 BTV 业务，建议三层多播 IPv6 地址使用临时多播地址，多播范围限定为组织本地范围，即：属于组织的多个站点。

IPv6 多播地址有 32bit 的多播组 ID，可以根据实际需要灵活使用。

（5）IPv6 地址规划示例

IPv6 地址长度为 128bit，IP 地址可以携带丰富的管理信息，在进行 IPv6 地址规划的时候，尽量通过地址规划简化管理。

图 5-32 是某运营商 IPv6 地址规划方案。

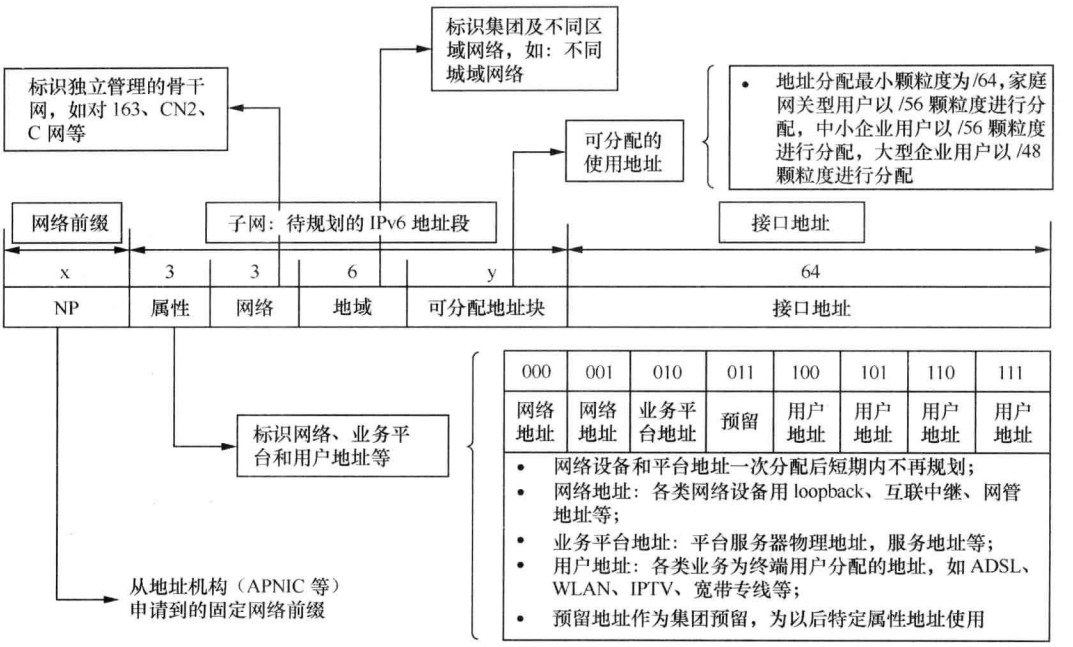


图 5-32　某运营商 IPv6 地址规划方案

注意：把业务属性放在网络 + 地域前面，这样是为了便于基于地址类型（设备地址、用户地址、业务平台地址）对路由进行控制。

ISP 总共可分配的 IPv6 地址数目可以通过如下公式计算出来（不考虑无效地址）：

$$2^{3+3+6+y} 或 2^{64-x}$$

如果 ISP 从 APNIC 申请到一个/32 的前缀，则其可以支配的 IPv6 地址总数为：$2^{64-32}=2^{32}$；

对于某个特定骨干网的某一地域，可分配的 IPv6 地址数目为：$2^{32}/2^{3+6}=2^{23}$；

对于某个特定骨干网某一地域，其可以分配的用户地址数目为：$2^{23}\times(4/8)=2^{22}$。

5.5　网络安全及业务溯源

5.5.1　IPv6 网络安全

“缺乏安全性是互联网天生的弱点，这与是否采用 IPv6 关系不大。事实上，IPv6 并没有引入新的安全问题，反而由于 IPSec 的引入，以及发送设备采用永久性 IP 地址，而解决了网络层溯源难题，给网络安全提供了根本的解决途径，有望实现端到端安全性。”中国电信科技委主任韦乐平这样评价 IPv6 网络安全。

1．IPv6 协议设计的安全考虑

从协议的角度，IPv6 作为 IPv4 的下一代，与 IPv4 同属于网络层的传输协议。然而，协议上最核心、最本质的差别就是地址空间的扩大。由 IPv4 下的 32bit 地址空间变为 128bit 的地址空间，这正是 IPv6 被选作新网络的承载协议并逐渐商用部署的根本驱动力。

IPv6 拥有如此巨大的地址空间，甚至可以为每一粒沙子都分配一个 IP 地址。而 IPv4 网络的地址分配是不规则的，并且很多时候是一个地址被多台主机共用。使用 IPv6 之后，我们

能够将每个地址指定给一个责任体，就像给每个人一个身份证号，每辆车一个车牌号一样，每个地址都是唯一的。IPv6 的地址分配采用逐级、层次化的结构，这就使得追踪定位、攻击溯源有了很大的改善。

另外，IPv6 提出了新的地址生成方式——密码生成地址。密码生成地址与公私钥对绑定，保证地址不能被他人伪造。这如同汽车的车牌印上了指纹，别人不可能伪造这样的车牌，因为指纹造不了假。

在 IPv6 协议设计之初，IPSec（IP Security）协议族中的 AH（Authentication Header，报文认证头）和 ESP（Encapsulation Security Payload，报文封装安全载荷）就内嵌到协议栈中，作为 IPv6 的扩展头出现在 IP 报文中，提供完整性、保密性和源认证保护，这无疑是从协议上较大地提升安全性。

整体上看，IPv4 协议的设计没有任何的安全考虑，特别是报文地址的伪造与欺骗使得无法对网络进行有效的监管和控制。因此，当出现网络攻击与安全威胁时，我们只能围绕攻击事件做好事前、事中和事后的防范、检测和过滤防御，缺乏有效的技术支撑手段，无法对攻击者形成真正的打击和管控。

而在 IPv6 网络的安全体系下，用户、报文和攻击可以一一对应。用户对自己的任何行为都必须负责，具有不可否认性。所以 IPv6 建立起严密的围绕攻击者的管控机制，实现对用户行为的安全监控。

2．IPv6 能减缓现有攻击

扫描几乎是任何攻击手段的必需前提。攻击者利用扫描收集目标网络的数据，据此分析、推断目标网络的拓扑结构、开放的服务、知名端口等有用信息，以作为真正攻击的基础。扫描的主要目的是通过 ping 每个地址，找到作为潜在攻击目标的在线主机或设备。

在 IPv6 时代，每个地址为 128bit，协议中规定的默认网络前缀为 64bit。换句话说，就是一个网段内有 264 个地址，假设攻击者以 10Mbit/s 的速度来扫描，也得需要大约 5 万年的时间才能遍历。IPv6 大大增大了扫描难度，由此增加了网络攻击的成本和代价。此时，黑客如果想侵占一定数量的主机发起 DDoS（Distributed Denial of Service，分布式拒绝服务）攻击，那么其将会付出更多的代价，这在一定程度上减少了 DDoS 攻击发生的可能性。

IPv6 协议定义了多播地址类型，而取消了 IPv4 下的广播地址，有效避免 IPv4 网络中的利用广播地址发起的广播风暴攻击和 DDoS 攻击。同时，IPv6 协议规定了不允许向使用多播地址的报文回复 ICMPv6（Internet Control Management Protocol version 6）差错报文，因此也能防止 ICMPv6 报文造成的放大攻击。

IPv6 下的密码生成地址是新的地址生成方式，它将公私钥对中的公钥与 IPv6 地址进行绑定。使用此类地址，能够保证报文的源地址不被他人伪造。在这样的安全机制保护下，在网络中传输的每一个报文均对应于一台主机，如果发生任何的攻击或者违法犯罪行为，都能够根据攻击报文追踪到发出此报文的主机，进而追查到攻击者。这种可靠的追踪溯源机制，使得黑客和攻击者容易被发现，这样就减少了网络攻击发生的可能性。

3．IPv6 面临的新威胁

IPv6 协议对 IPv4 协议的根本改变发生在 IP 层，因此针对 IPv6 协议所定义的分组头及扩展头的容易发生的安全威胁，我们需要进行充分的准备。常见的针对 IPv6 扩展头的攻击，

主要包括利用分片扩展头发起分片攻击；逃避防火墙或 IDS（Intrusion Detection System，入侵检测系统）的检查或者发动 DDoS 攻击；利用路由扩展头的 type 0 类型，在网络中发起放大攻击。

在 IPv4 向 IPv6 的演进过程中，我们还需要考虑各种过渡技术与方案的安全隐患。由于在共存时期，IPv4 网络与 IPv6 网络同时存在，且有互通需求，这就要求来自两种不同 IP 协议网络的威胁不能够交叉感染。另外，运营商所选择部署的各种过渡技术，由于尚无成熟的使用经验，因此很可能存在潜在的安全风险。

具体来看，隧道技术是对报文进行一层封装。隧道报文在经过网络安全设备时，如果需要被检查与过滤，那网络安全技术就得支持各种新的隧道协议，能够对隧道进行解封装，而后才能对内嵌报文进行处理。在建立各种隧道的时候，对隧道节点的认证是十分必要的。否则，所建立的隧道是不可靠的，隧道就会被黑客和攻击者利用，成为进入某个目标网络的通路。

而在利用翻译技术将 IPv6 与 IPv4 网络进行互联互通时，要改变报文的 IP 层及传输层的相关信息，这样会对端到端的安全产生影响，导致 IPSec 的三层安全隧道在翻译设备处出现断点。同时，翻译设备作为网络互通的关键节点，会成为 DDoS 攻击的主要目标。

总之，安全隐患的主要根源是网络协议设计的缺陷和网络设备开发的 bug，或者是网络协议的部署与使用出现问题。在 IPv6 商用之初，我们应利用 IPv6 协议提供的新安全特性，先解决部分类似 IPv4 下已有的威胁。随着 IPv6 网络使用范围的扩大以及各种应用的迁移和增多，我们需要更加关注新型攻击。

4．IPv6 的安全机制和策略

（1）充分利用地址自身的安全特性

在合适的场景下，使用密码生成地址有助于发现针对 ND（Neighbor Discovery，邻居发现）协议和 DHCPv6（Dynamic Host Configuration Protocol for IPv6，IPv6 动态主机配置协议）协议的欺骗和伪造等问题。由于拥有自身的公私钥对，密码生成地址还可用于 IPSec 的协商，在某些场景下可以简化协商过程，提高 IPSec 的性能。

另一种关于地址的安全机制是隐私扩展。由于 IPv6 网络不需要地址转换，因此没有 NAT（Network Address Translation，网络地址转换）设备，这样一来就会将内网的结构及相关信息暴露。所以，内网的网络设备要依靠隐私扩展机制，通过周期性的地址改变，防止内部网络信息的泄露。

（2）设定精细的过滤策略

面对 IPv6 地址结构以及相关协议的改变，防火墙或者网络边界设备需要设定更加精细的过滤规则。

防火墙需要拒绝对内网知名多播地址访问的报文，关闭不必要的服务端口，过滤内网使用的地址。在对 ICMP 报文的处理上，由于 IPv6 对 ICMPv6 的依赖程度远远超过了 IPv4（ICMPv6 除了完成 ping 及错误消息外，还新增了地址分配、地址解析、多播管理和移动 IPv6 下的使用等功能），因此 ICMPv6 报文的过滤策略需要根据实际情况小心设置，避免影响正常的服务和应用。

为了防护 IPv6 扩展头的隐患，防火墙需检查扩展头的合法性。对于分片报文，防火墙能拒绝发送到网络中间设备的分片，并支持重组，具备防 DDoS 攻击能力。防火墙能识别 Type

0 类型的路由扩展头报文，并对其进行过滤。入口过滤机制在防火墙和边界设备上的实现也是必要的，这样可以缓解网络间的源地址伪造威胁。

（3）合理的部署策略

在 IPv6 与 IPv4 共存时期，过渡技术的安全部署尤为重要。各种网络安全设备都要具备对 IPv6 和 IPv4 威胁的防护能力。我们在使用隧道技术时，要尽可能采用静态配置隧道，以降低动态隧道的伪造和非法接入威胁。防火墙要设置对非授权隧道报文的过滤，同时识别各种隧道协议，能够对隧道报文的内嵌封装报文做访问控制。翻译设备则要做好自身的 DDoS 攻击防范。

2010 年是 IPv6 的商用元年，各大运营商纷纷制定过渡方案，宣布演进计划，开启试点城市。华为的安全产品已经为 IPv6 商用做好了充分准备，2009 年通过了 IPv6 Ready 认证。值得一提的是，华为 IPv6 防火墙在国内首家通过 IPv6 Ready 第二阶段增强认证，而且能将当前多种主流的过渡技术与方案集成于同一设备，满足了当前 IPv6 商用的市场需求。

5.5.2 IPv6 业务溯源

（1）需要用户溯源的原因

用户溯源是指根据源地址、端口等信息，确定最终用户的技术方式。

以国家安全监控为例。例如，某人在网络上发表了反动言论的帖子，网络中会记录某时间某用户发表了某言论，其用户信息为<用户的公网 IPv4 地址，用户公网端口号>。相关国家安全组织可以根据时间和公网 IPv4 地址的信息（Radius Server 记录所有用户的上下线时间和公网 IPv4 地址分配情况）检索到该用户。

（2）NAT 方案下溯源的复杂性

部署了 CGN 之后的用户溯源，不再是根据 IPv4 地址来识别用户，而是根据公网 IPv4 地址和端口号来识别用户。

由于用户占用公网 IPv4 地址和端口号的不确定性，也增加了用户溯源的复杂性。

① 打破 IP 地址和用户的对应关系。

② 需要根据（公网地址 + 端口号）才能对应到用户。

③ 溯源带来的海量 NAT 日志和端口块映射等问题，增加运维难度和成本。

5.5.3 运营级协议转换技术网络安全

1. 安全方案概述

端到端网络安全方案，包括管理安全、设备安全、用户安全和业务安全 4 个方面，见表 5-8。

表 5-8　　安全方案

安全类型	子类型	OLT	BRAS
用户安全	用户账号安全	PPPoE + 标识用户物理位置	基于用户账号和物理位置信息进行认证
	方案用户间二层攻击	Smart VLAN 隔离用户	VLAN 隔离用户

续表

安全类型	子类型	OLT	BRAS
业务安全	用户仿冒源 MAC 地址攻击	限制用户 MAC 地址学习数；防御 MAC-Spoofing 功能	用户侧接口配置为 untrusted 模式；限制用户 MAC 地址学习数
	用户仿冒源 IP 地址攻击	防御 IP-Spoofing 功能	防御 MAC/IP-Spoofing 功能
	DHCP Server IPv4 地址池耗尽攻击	—	Hops = 0，检查报文源 MAC 和 CHARR 是否匹配

2．管理安全

管理安全主要包括如下措施。

① 对设备开放的服务预先进行规划，避免开启不必要的服务带来攻击隐患。

② 对用户登录权限进行严格的控制和认证。禁止非法用户登录设备，结合 TACACS 控制权限等级。

3．设备安全

目前，接入设备对 IPv6 的设备安全性考虑还很少，本节内容主要聚焦于路由器设备的安全防护措施。

（1）CP-CAR

CP-CAR 用来设置上送 CPU 报文的分类限速规则，针对每类报文可设置均值速率、承诺突发尺寸、优先级信息等。通过对不同的报文设置不同的 CAR 规则，可以降低报文的相互影响，达到保护 CPU 的目的。CP-CAR 还可以设置上送 CPU 报文的整体速率。当整体上送速率超过阈值时，报文将被丢弃，以避免 CPU 过载。

CP-CAR 一般与白名单和黑名单配合使用。白名单指合法用户或者是高优先级用户的集合，匹配白名单特征的报文采用高速率高优先级上送；黑名单指非法用户的集合，匹配黑名单特征的报文被丢弃或者低优先级上送。

（2）应用层联动

应用层联动指的是将控制层面的协议开关状态和底层转发引擎的协议上送关联起来。通过在上层和底层建立联系，在协议开关状态上保持一致。对于设备没有开启业务的协议，底层硬件默认以小带宽上送其协议报文，也可以配置为完全不上送，这样就将攻击者的攻击范围尽可能缩小，增加了攻击的成本，减小了设备的安全风险。

（3）路由协议安全

BGP4 + 、IS-ISv6 支持 MD5 认证，OSPFv3 支持 IPSec 认证。

4．用户安全

用户安全指的是为保障个人用户业务安全使用而提供的安全防护措施，如图 5-33 所示。

（1）用户账号安全

通过把用户物理位置信息和用户账号信息进行绑定认证，避免恶意用户盗用行为。具体措施如下。

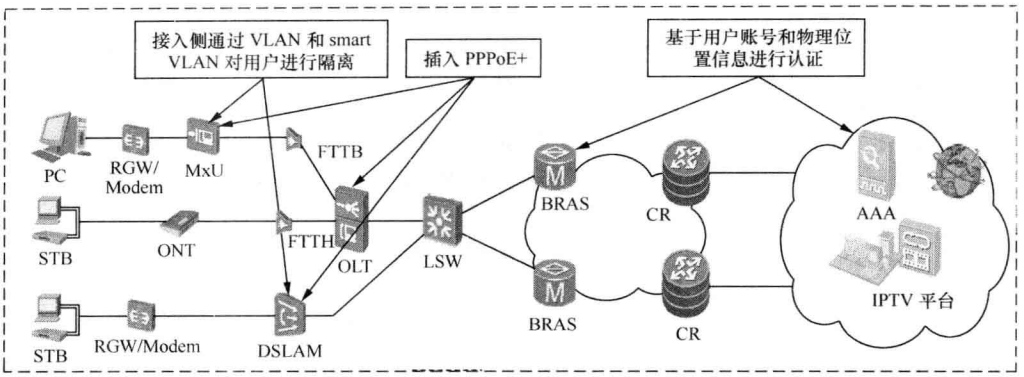


图 5-33　用户安全部署

① BRAS 和 AAA 通过 CHAP 协议进行认证。

② 接入设备插入 PPPoE + 信息，携带用户接入端口信息。

③ AAA 通过绑定用户名/密码、接入端口、BRAS IP 等信息精确标识用户。

（2）用户端口二层隔离

通过 PVC/VLAN 或 smart VLAN 实现用户端口的二层隔离，防范用户二层攻击（如恶意用户仿冒 DHCP Server 等二层攻击行为）。

5．业务安全方案

业务安全指的是保障大量用户业务安全使用而提供的安全防护措施。不同于用户安全方案，业务安全聚焦在保护所有用户共用资源的安全，如图 5-34 所示。

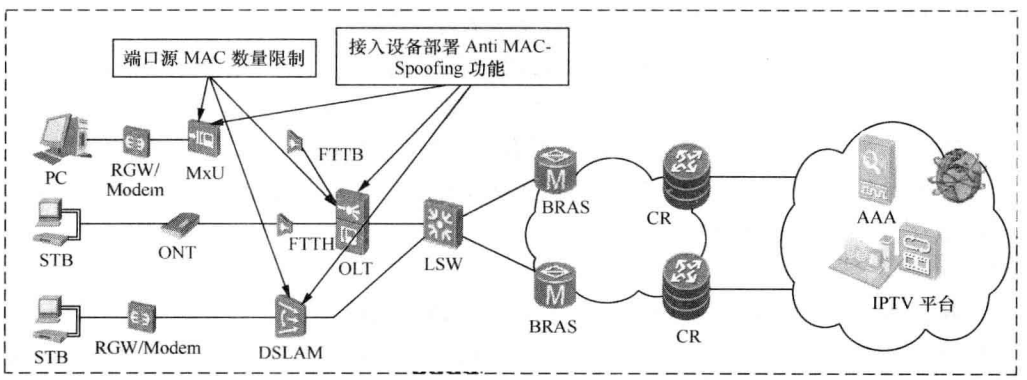


图 5-34　业务安全部署

（1）防范 MAC 欺骗和非法用户报文

接入网设备 MxU/OLT 部署 Anti MAC-Spoofing 功能。认证通过时，系统根据用户发送的 PPPoE 报文将 MAC 地址和用户端口动态绑定，最大绑定 8 个 MAC 地址。只有业务流的源 MAC 地址与绑定的 MAC 地址一致时，才允许其通过设备上行，否则丢弃，从而防止接入用户通过 MAC 地址欺骗方式对设备进行攻击。

（2）防范 IP-Spoofing 攻击

BRAS 支持 DHCPv4 Snooping 功能。通过 Snooping 建立 DHCP 绑定表，将 MAC、VLAN、端口和 IP 地址绑定，防范恶意用户伪造源 IP 的攻击。

5.5.4　运营级协议转换技术业务溯源

1. CGN 的引入增加了溯源的复杂性

运营商级 NAT（Carrier-Grade NAT，CGN）又称作大规模部署 NAT（Large-Scale NAT，LSN）。与普通 NAT 相比，CGN 主要在支持并发用户数、性能、溯源等方面有很大提高，以适应运营商的大规模商业部署，快速解决 IPv4 地址短缺的急迫问题。CGN 有多个应用场景，包含了前面我们所说的 NAT444、DS-Lite、6RD 等。

CGN 的部署面临自身的一些问题，如地址转换增加了报文传输时延，用户溯源难，某些应用穿越 NAT 难等。在 CGN 的开发部署中，这些问题被作为重点逐个解决了。CGN 相对于传统的 NAT 业务，还有诸多好处，比如端口预部署，动态溯源等，使得 CGN 更适合运营级解决方案。

部署 CGN 后增加了溯源的难度。用户报文在 NAT 转换前后的源地址不同，安全监管机构只能获得 NAT 转换后的用户信息。这个地址在 AAA 没有任何记录，也没有与用户信息关联，因此无法完成正常的用户信息反查和用户溯源，如图 5-35 所示。

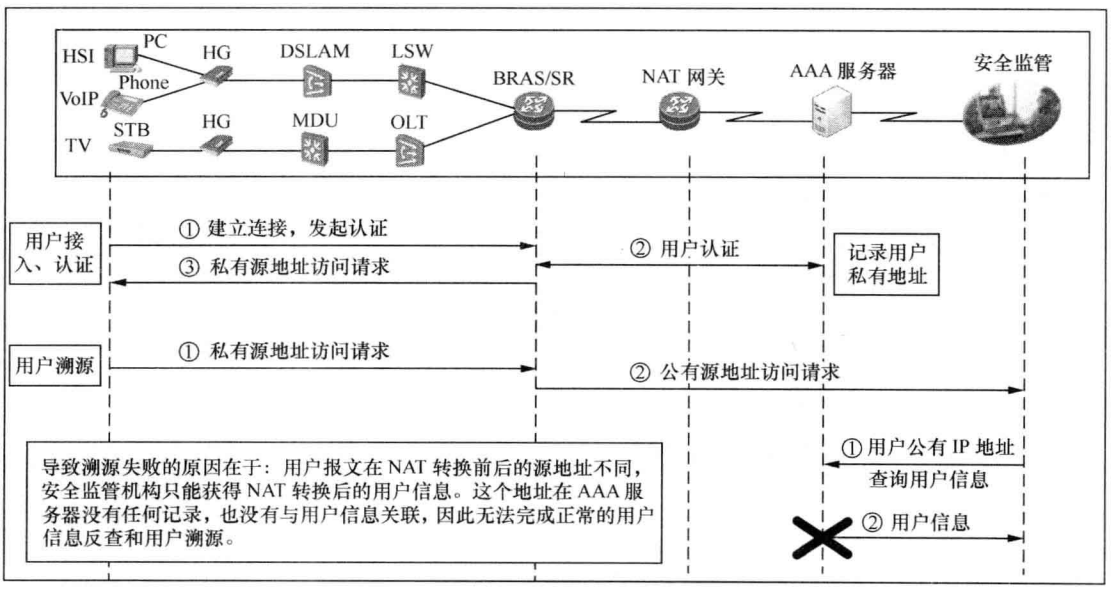


图 5-35　用户信息溯源失败

部署 CGN 之后的溯源方式如下。

① 动态溯源：在线溯源，又称为动态映射。

② 离线溯源：syslog 方式溯源。

2. 端口预部署在溯源方案中的应用价值

根据经验值可知，一个主机可能会消耗 200 个端口号。在没有端口号预分配的情况下，CGN 设备会根据内部流量情况随机给主机分配公网 IPv4 地址和端口号，导致一个主机可能会占用多个公网 IPv4 地址。

假设一个 CPE 下有 10 台主机，那么一个 CPE 可能会消耗 2 000 个端口号，占用的公网

IPv4 地址可能会更多。

端口预部署的本质就是为 CPE 分配一个固定公网 IPv4 地址和端口号范围。这样，该用户所有的 NAT 信息都会集中在该公网 IPv4 地址和端口号范围之内，而且 Radius Server 记录所有用户的上下线时间和私网 IPv4 地址、公网 IPv4 地址和公网端口号范围信息。这样，用户溯源就会变得简单。

3. 动态溯源方案

动态溯源也称为 BRAS 上报映射表，如图 5-36 所示。该方案仅适用于 BRAS 插卡的 CGN 部署方式。该方案由 BRAS 生成用户地址映射关系，并上报 AAA。

① BRAS 为用户地址选择公有地址、端口块，创建用户地址映射关系，保证为不同用户地址选择不同的公有地址及端口块。

② BRAS 通过扩展 Radius 属性，在 accounting-Request 消息中上报用户地址对应的公有地址、端口块等信息。

③ AAA 获得用户地址、公有地址、端口块等信息，并维持与用户信息的对应关系。

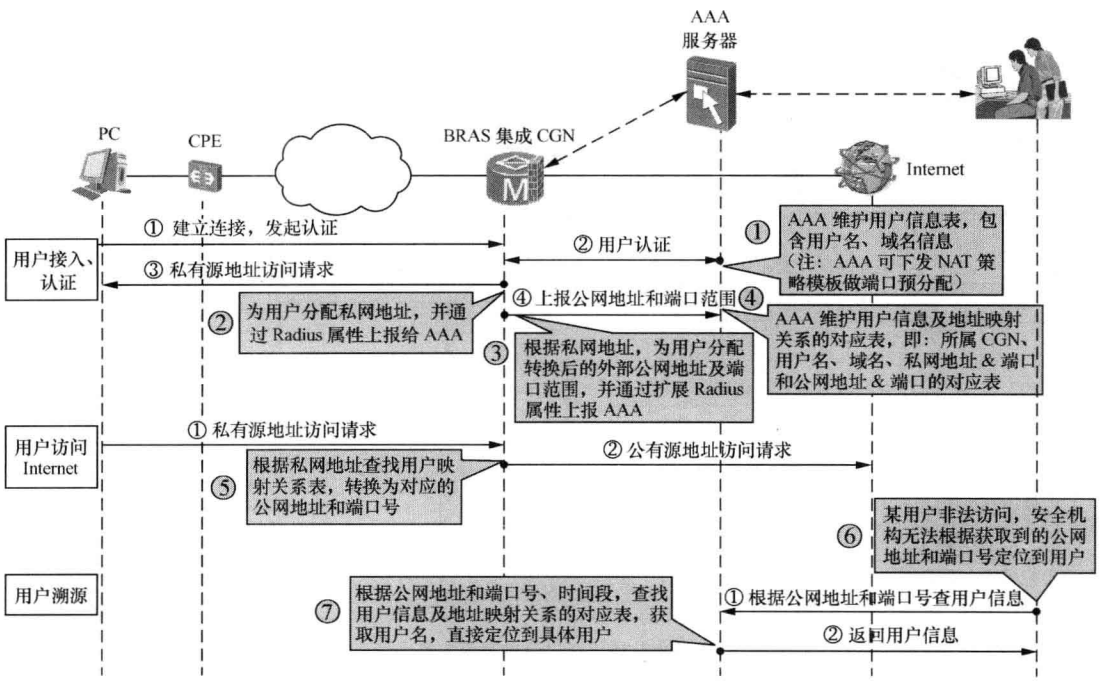


图 5-36　动态溯源方案

4. 离线溯源方案

在离线情况下，安全机构通过查询 Log 服务器和 AAA 服务器，获取用户信息。该方案适用于插卡式 CGN（如插卡在 CR 或 BRAS）及独立式 CGN 的所有 CGN 部署。通常 Log server 会储存至少 3～6 个月的用户日志，如图 5-37 所示。

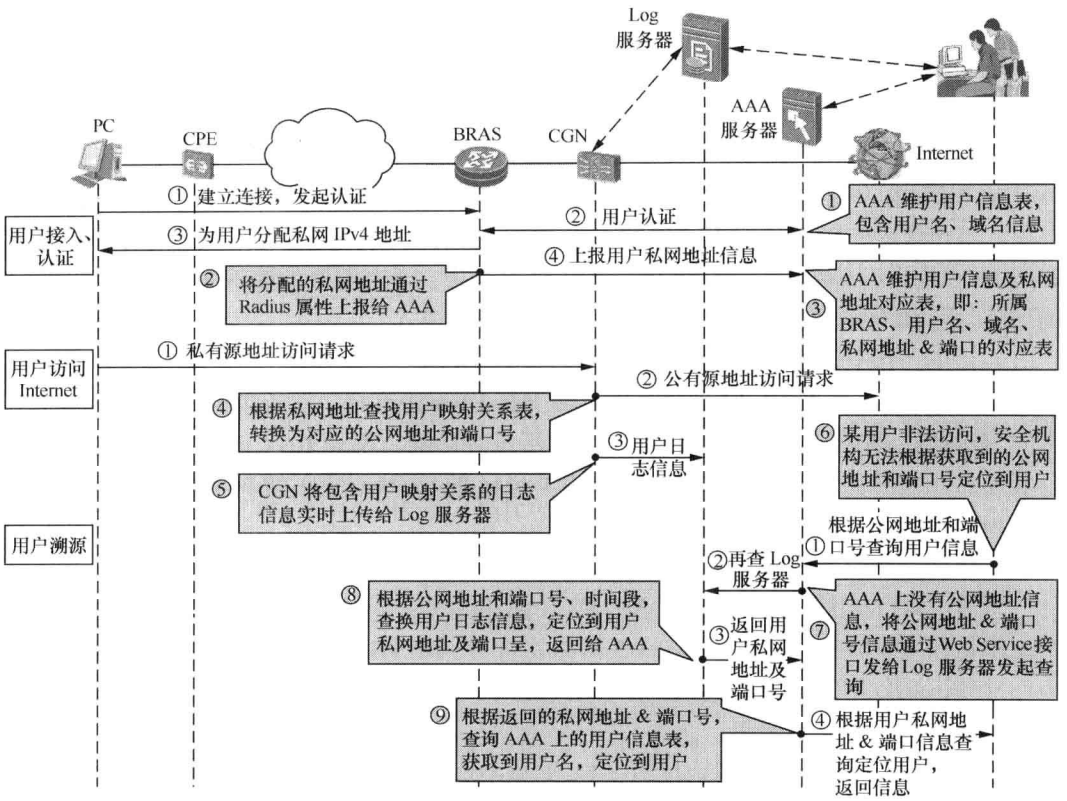


图 5-37　离线溯源方案

5.6　总体演进策略建议

由于 IPv6 与 IPv4 的不兼容性，使得现有互联网从 IPv4 向 IPv6 的演进不可能在同一时刻全部完成，也因此引发了关于 IPv6 演进阶段性策略的讨论。对于运营商而言，如何保护现有 IPv4 网络大量的投资，同时又尽量减少对用户和应用的影响，也是在完成网络整体演进过程中需要关注的主要问题。为了保证网络向下一代互联网的平滑演进，IPv4 向 IPv6 过渡时必须考虑以下 4 个原则。

① 业务驱动。下一代互联网演进的主要驱动力来自业务，一方面需要保证现有业务的发展不受 IPv4 地址的局限，另一方面也要支持未来新兴业务（如三网融合、物联网应用）的发展需求。

② 过渡平稳。保证过渡阶段现有网络稳定运行，IPv4 的用户和业务体验尽量不受影响。过渡技术和组网方案应减少对网络架构和系统架构的影响，确保过渡期的平稳性。

③ 技术创新。在过渡的不同阶段会面临不同的场景和需求，应关注技术创新和新技术的部署实施，关注重点技术，加快翻译、隧道技术的成熟商用。

④ 成本兼顾。过渡方案在满足业务需求和网络演进的基础上，应尽可能兼顾网络部署的难度和成本，降低网络部署的复杂度，确保可持续发展。

总体来说，IPv6 演进是庞大的系统工程，在时间跨度、技术复杂性，以及资金投入上都不能掉以轻心，需要全面慎重谋划，实施尽可能平滑的过渡演进。演进策略关键是使用户和

互联网 SP/CP 能够以合理的成本，使用 IPv6 协议和规则接入 IPv6 互联网，在更大的网络平台上进行安全、可靠、可管可控的通信，同时还要始终与现有 IPv4 网实现互联互通，能够使用现有的 IPv4 业务和应用。

IPv6 过渡总体演进策略需要考虑多种场景下 IPv6 过渡技术的引入和部署，包括 IPv4 和 IPv6 主机和应用的共存，以及 IPv4 和 IPv6 网络互通的问题。基于 IPv4 的互联网已经获得蓬勃发展，如此众多的用户和应用迁移到支持 IPv6 网络的周期较长，而种类繁杂的网络设备升级支持 IPv6 栈也有待时日。在如此漫长的迁移过程中，现有网络的演进需要考虑不同阶段业务的迁移程度和网络升级改造程度，采用不同的演进模式。

（1）初期：现有 IPv4 网络部署 CGN 解决地址问题

现有网络以 IPv4 为主，网络设备仍主要支持 IPv4 协议栈。对于现有的 IPv4 应用，仍由原有的 IPv4 网络和业务平台提供，对用户使用没有影响。考虑 IPv4 地址枯竭问题，对于 IP 地址紧缺的区域和网络，通过部署支持 NAT44/444 的 CGN 设备，解决现有 IPv4 网络的私网地址使用问题，缓解 IP 地址压力。

（2）中期：网络采用双栈 + NAT 实现 IPv4 和 IPv6 业务共存

该阶段 IPv6 应用已经开始形成小规模，IPv4 和 IPv6 流量并存，局部网络已支持 IPv6 栈。骨干网络支持 IPv4/IPv6 双栈，部分城域网络支持 IPv4/IPv6 双栈，但并存的 IPv4 网络仍需使用公网地址，因此需要基于双栈网络通过地址翻译技术解决双栈部署中 IPv4 主机和应用采用私网地址的问题。

对于 IPv4 公网地址和私网地址转换技术（也统称为 NAT 技术），运营商通常部署 CGN（运营商级 NAT）设备支持多种 NAT 技术（NAT44、DS-Lite）。其与传统 NAT 的差异在于，该设备可提供对大量运营商用户的转换，具备更高的会话处理能力和高可靠性。通过双栈 + NAT 的部署，将可以解决 IPv4 和 IPv6 从主机到网络，再到应用的共存，同时以私网解决 IP 地址短缺问题。

（3）中后期：网络全面升级为 IPv6，通过隧道实现 IPv4 业务穿透

IPv6 演进中后期，IPv6 用户和应用已成为主流，网络中所有设备已完全支持 IPv6 栈。在此阶段，网络可以全面支持双栈，同时支持端到端的 IPv4 和 IPv6 应用传送。网络也可能关闭 IPv4 栈，而只保留 IPv6 栈，以降低同时运行双栈的维护复杂度和运维成本。此时网络主要传送 IPv6 数据分组，而对于少量没有迁移到 IPv6 的 IPv4 用户和应用，则通过 IPv6 隧道封装 IPv4 数据分组进行传送，例如 DS-Lite 等过渡技术。等到 IPv4 终端和操作系统全面实现向 IPv6 的演进的时候，从网络到终端全面 IPv6 化，将通过 NAT64 部署实现 IPv4 和 IPv6 网络的互通。

对于运营商的网络而言，由于 IPv6 地址与 IPv4 地址的不兼容，从现有 IPv4 网络向 IPv6 网络过渡将持续较长一段时间。在下一代互联网的演进过程中，必须遵循解决现有网络问题、坚持平滑过渡和因地制宜的原则，才能高效实现网络、业务和用户的顺利升级过渡。因此，下一代互联网是一个不断发展中的整体架构，需要多个体系的协同演进和推进。网络运营商可基于不同的业务驱动需求及网络能力现状，选择不同的演进模式和部署策略，包括新建模式、端到端双栈模式和阶段过渡模式。这其中，过渡技术已经成为下一代互联网演进的关键要素，有待进一步深入研究和实践，以加快过渡技术的成熟和规模部署。

另一方面，在 IPv4 向 IPv6 的过渡过程中，网络运营商首先会面临 IPv4 地址短缺问题，

因而也有最为迫切的 IPv6 引入和过渡需求。但是，IPv6 的演进又是一个系统性的复杂工程，涉及网络设备、终端设备、操作系统软件、内容应用、技术研发等多个环节，需要涉及的各个环节联合运作，才能推动现有互联网体系向下一代互联网平稳快速过渡。

基于各过渡技术的进展，运营商可以根据自己的演进特点，选择合适的过渡方案组合。从目前国内的实践看，IPv6 引入进度较快的省份，大多是以双栈网络部署为基础来引入 IPv6 应用。双栈网络准备好后，即可以 IPv6 网络的逻辑平面能力来孵化 IPv6 业务和应用。在全网具备了双栈能力的基础上，通过 NAT44 和 NAT64 等翻译技术解决地址紧缺问题。与此同时，部分省份也在试验 DS-Lite 等纯 IPv6 网络的过渡方案。

第 6 章
内容和应用过渡技术部署

IPv6 不仅仅是一个技术问题，更是一个商业问题，而商业问题上的产业链协同发展往往更为困难。因此，IPv6 产业链仅仅依靠某一个方面推动是很难走向成熟的，这需要产业链各方协同创新与努力。

抢占 IPv6 的先机，不仅意味着抢占了一项技术、一个标准，更遥远的意义在于，在万物都在联网的将来，IPv6 将联通万物，各行各业都要牵涉其中，带来的商机是无限的。从基础网络到通信运营商，再到互联网服务商乃至用户，IPv6 过渡是一项庞大的系统工程，涉及运营商、ICP、网络设备提供商、终端设备、软件、应用等多个环节，必须全体联动，才能顺利完成。任一环节存在短板，都可能使 IPv6 发展停滞不前。

虽然我国已经明确了从 IPv4 向 IPv6 迁移的路径以及策略，但是受制于目前国内互联网产业发展以及相关体制的不完善，IPv6 的推进进程还相对缓慢。这首先在于 IPv6 与现有 IPv4 不兼容，网络和网站运营商必须升级他们的网络设施和软件以支持 IPv6 信息，但升级转换牵涉方方面面，尤其是软件应用全面转换升级难度较大。其次，IPv4 向 IPv6 过渡、两者并存的方案并不够清晰。目前，在部署 IPv6 方面，运营商、互联网商、通信设备商仍然各自为战。此外，应用的匮乏是面临的最大难题之一。

综上所述，现网向 IPv6 过渡的成功与否很大程度上取决于内容和应用能否顺利迁移，而内容和应用迁移的重点是 IDC 升级改造、ICP 内容和应用迁移、物联网等新兴应用的迁移。

6.1 IDC 的双栈升级改造方案

IDC 即是 Internet Data Center，是基于 Internet 网络，为集中式收集、存储、处理和发送数据的设备提供运行维护的设施基地，并提供相关的服务。IDC 提供的主要业务包括主机托管（机位、机架、机房出租）、资源出租（如虚拟主机业务、数据存储服务）、系统维护（系统配置、数据备份、故障排除服务）、管理服务（如带宽管理、流量分析、负载均衡、入侵检测、系统漏洞诊断），以及其他支撑、运行服务等。

如上所述，IDC 的双栈化改造是实现内容和应用向 IPv6 迁移的基础和前提。IDC 基础网络的改造主要通过升级或替换原有的设备和系统（包括出口路由器、汇聚/接入交换机、防火墙和网管系统等），支持 IPv4/IPv6 双栈功能，承载 CP/SP 的 IPv6 数据，并与骨干网和城域网 IPv6 的商用部署相结合，支持 CP/SP 改造，实现用户对 IDC IPv6 业务的访问，同时

实现 IDC IPv6 业务的开放与配置、网络系统安全管理、QoS 流量控制、网络管理等生产运维行为。

6.2　典型内容和应用迁移方案分析

目前典型的互联网应用主要包含以下几个方面：网页浏览、FTP、电子邮件、即时通信、流媒体等应用。以上几种应用迁移的方案建议如下。

6.2.1　网页浏览类应用迁移方案

网页浏览所使用的协议主要包括 HTTP、HTTPS 协议，因此需详细分析以上两种协议的工作原理和过程，以分析 IPv6 引入后对以上两种协议的影响。

1．IPv6 对网页浏览应用的影响分析

（1）对 HTTP 协议的影响

要了解 IPv6 对网页浏览协议（即 http 以及 https 协议）的影响，首先要了解 http/https 协议的工作原理。用户使用 http 协议上网的原理如图 6-1 所示。

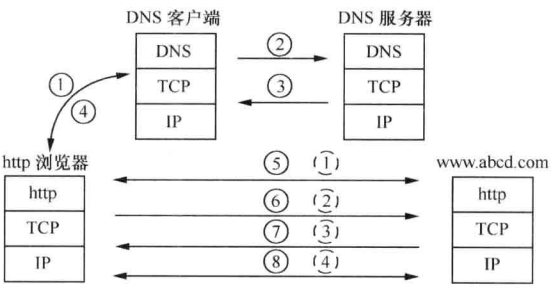


图 6-1　用户使用 http 协议上网原理图

用户访问互联网网页存在两种形式，一种是通过域名访问，另一种是直接通过 IP 地址访问。通过域名访问的流程见图 6-1 中实线圆圈所标识的序号，通过 IP 地址访问的流程见图 6-1 中用虚线圆圈所标识的序号。

对于用户使用域名访问的情况，用户在浏览器中键入 www.abcd.com，浏览器则自动将其转换成 http://www.abcd.com/。该 URL 是这样组成的，http://代表超文本传输协议，通知 abcd.com 服务器显示 Web 页，通常不用输入；www 代表一个 Web（万维网）服务器；abcd.com 这是装有网页的服务器的域名，或站点服务器的名称。

① 浏览器从 URL 中抽取出主机名 abcd.com 后，把它传递给本地主机上的 DNS 应用客户端。

② DNS 客户端发送 DNS 请求报文到 PC 配置好的 DNS 服务器。该请求可以是申请 IPv4 地址的 A 请求，也可以是申请 IPv6 地址的 AAAA 或者 A6 请求。

③ DNS 服务器通过域名解析，返回给 DNS 客户端该域名所对应的 IP 地址，根据申请的类型返回 IPv4 或者 IPv6 地址。

④ DNS 客户端将该 IP 地址返回给浏览器。

⑤ 浏览器接着打开一个到位于该 IP 地址的 HTTP 服务器的 TCP 连接。

⑥ TCP 连接建立后，由浏览器发起一个 http 请求报文。

⑦ 相应的 HTTP 服务器返回一个 http 响应报文。

⑧ 用户进行页面浏览，当用户关闭网页，浏览器发起关闭请求，关闭 TCP 连接。

http 请求报文由 3 部分组成，分别是：请求行、消息报头、请求正文，具体如下。

请求行以一个方法符号开头，以空格分开，后面跟着请求的 URI 和协议的版本，格式如为 Method Request-URI HTTP-Version CRLF。其中，Method 表示请求方法，包括 GET、POST、HEAD 等；Request-URI 是一个统一资源标识符；HTTP-Version 表示请求的 HTTP 协议版本；CRLF 表示回车和换行。例如，GET 方法：以在浏览器的地址栏中输入网址的方式访问网页时，浏览器采用 GET 方法向服务器获取资源，具体表示为 GET/form.html HTTP/1.1（CRLF）。

http 报文报头是可选的，包括普通报头、请求报头、回应报头以及实体报头，这里只描述请求报头。请求报头用于客户端向服务器端传递请求的附加信息以及客户端自身的信息。请求报头包括 Accepet、Authorization、Host 以及 User-Agent 报头，其中 Host 报头是必须的。Host 请求报头域主要用于指定被请求资源的 Internet 主机和端口号，它通常是从 HTTP URL 中提取出来的，例如：我们在浏览器中输入 http://www.abcd.com，浏览器发送的请求消息中，就会包含 Host 请求报头域，Host：www.abcd.com，此处使用缺省端口号 80，若指定了端口号，则变成 Host：www.abcd.com:指定端口号。

请求和响应消息都可以传送一个实体。一个实体由实体报头域和实体正文组成，但并不是说实体报头域和实体正文要在一起发送，可以只发送实体报头域。实体报头定义了关于实体正文（有无实体正文）和请求所标识的资源的元信息。实体正文就是返回的网页内容代码。

HTTP 响应也由 3 个部分组成，分别是状态行、消息报头、响应正文。

状态行：HTTP-Version Status-Code Reason-Phrase CRLF，其中，HTTP-Version 表示服务器 HTTP 协议的版本；Status-Code 表示服务器发回的响应状态代码；Reason-Phrase 表示状态代码的文本描述。

状态代码由 3 位数字组成，第一个数字定义了响应的类别，且有 1xx～5xx 5 种可能取值。例如，HTTP/1.1 200 OK（CRLF）表示请求已被成功接收、理解、接受。

响应报头：允许服务器传递不能放在状态行中的附加响应信息，以及关于服务器的信息和对 Request-URI 所标识的资源进行下一步访问的信息。例如，Location 响应报头域用于重定向接收者到一个新的位置。Location 响应报头域常用在更换域名的时候。Server 响应报头域包含了服务器用来处理请求的软件信息，与 User-Agent 请求报头域是相对应的。WWW-Authenticate 响应报头域必须被包含在 401（未授权的）响应消息中。客户端收到 401 响应消息，并发送 Authorization 报头域请求服务器对其进行验证时，服务端响应报头就包含该报头域。例如，WWW-Authenticate:Basic realm = "Basic Auth Test!"，可以看出服务器对请求资源采用的是基本验证机制。

以上是用户采用域名的方式访问互联网的流程，如果是用户在浏览器内直接键入 IP 地址，或者直接点击一个包含 IP 地址的 URL 时，则不需要从 DNS 处进行域名解析，直接从 http 浏览器到目的服务器发起 TCP 连接直到最后关闭网页，流程同⑤、⑥、⑦和⑧。

通过对用户访问网页流程的分析可以发现，当 IPv6 引入后，用户是否能够成功访问网页，与 HTTP 协议本身并没有关系，而是与客户端浏览器对底层 IP 协议的调用以及 URL 中的域名或者 IP 地址有关。同时，相关联地，与该被访问的网站以及用户终端所支持的协议有关。关系如图 6-2 所示。

对于使用域名访问互联网的用户，如果终端是单栈，则只能发起 A 或者 AAAA 的请求，返回 IPv4 或者 IPv6 地址，对于单栈的网站，则存在无法返回用户所需类型地址的情况。例如，一

个未迁移到 IPv6 的 IPv4 则无法返回 IPv6 单栈用户所需的 IPv6 地址，导致用户访问失败。

对于使用 IP 地址访问互联网的用户，情况则更加复杂一些。由于 http 服务器返回给 http 浏览器客户端的网页中所包含的链接是以 IP 地址的形式呈现的，因此即便该链接所对应的网站已经改成了双栈，此 IP 地址还是只可能将用户引向一个单栈的网页。例如，该首页中引用的网页是进行过双栈改造的，但是 URL 中只写入了该链接的一个 IPv4 地址，这样对于一个 IPv6 单栈的用户，就无法访问该网页。但事实上，如果 URL 中采用的是域名的形式，该 IPv6 用户是可以访问成功的。

（2）对 HTTPs 协议的影响

https 并不是一个单独的协议，可以把它理解成将一个普通的 http 协议运行在加密的 SSL 层或者 TLS 层上。其中，加密的 SSL（encrypted Secure Sockets Layer）是运行在应用层之下传输层之上的一个子层，TLS（Transport Layer Security）是运行在传输层的一种传输层安全协议。

用户使用 https 来访问网页时的流程如图 6-3 所示。

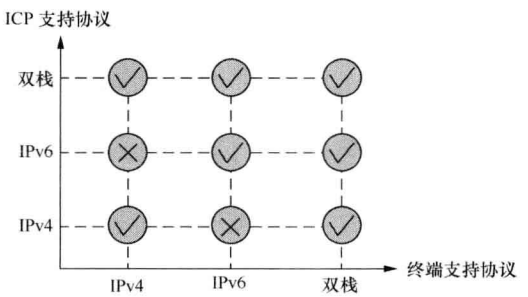


图 6-2　影响用户访问网页成功因素

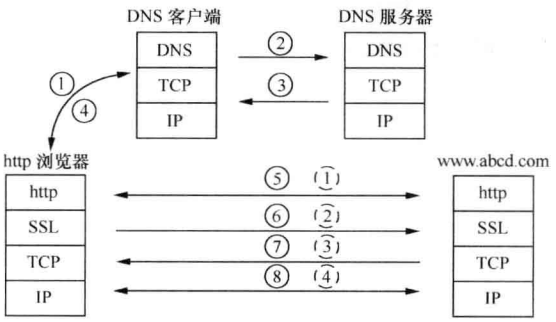


图 6-3　用户使用 https 协议访问网页流程图

与 http 访问流程唯一不同的地方在于第 5 步。SSL 层借助下层协议的信道，安全地协商出一份加密密钥，并用此密钥来加密 HTTP 请求。TCP 层与 Web Server 的 443 端口建立连接，传递 SSL 处理后的数据。接收端与此过程相反。SSL 在 TCP 之上建立了一个加密通道，通过这一层的数据经过了加密，因此可以达到保密的效果。

IPv6 对 https 的影响同 http，与 https 协议本身没有关系，而是与 URL 中的域名或者 IP 地址有关。同时，相关联地与该被访问的网站以及用户终端所支持的协议有关。

2．网页类浏览应用迁移方案建议

如前所述，IPv6 的引入对网页浏览的影响与 HTTP 协议本身没有关系，而是与 URL 中的域名以及 IP 地址相关。对于采用域名访问的用户，有 2 种类型的解决方案。一种是采用协议转换或者代理的方式，将 IPv4 的报文翻译成 IPv6 报文，从而给 IPv6 用户提供 IPv6 信息源；另外一种是采用双栈的方式给用户提供 IPv6 信息源。前者比较适合于在运营商的网络边界统一部署协议转换或者代理网关；后者比较适合于 ICP 内部自行改造。下面就这两种类型的方案进行具体阐述。

对于直接使用 IPv4 地址访问网页的用户，如果该用户是双栈用户，则可以直接用 IPv4 地址访问资源；如果是 IPv6 单栈用户，则无法访问该网站。

（1）运营商协议转换解决方案

此处的协议转换是通过部署 DNS ALG 来实现的。典型的技术是 NAT-PT 或者 NAT64。由于

用户在同一台终端上只能配置一个 DNS 地址，因此要将此 DNS 地址指向一个特定的 DNS ALG 地址，该 DNS ALG 只能由运营商统一部署在 IPv4 和 IPv6 网络边界并下发给用户。流程如下。

① 双栈用户 IE 向 IPv6 DNS 发起 AAAA 请求。

② IPv6 DNS 发现 ICP1 域名所对应的 IPv6 地址，该地址是 ICP1 的 IPv4 地址加上/96 的 prefix 形成的一个新的 IPv6 地址，手工配置在 DNS 里，返回给用户。

③ 双栈用户向 IPv4 DNS 发起 A 请求（与①同时）。

④ IPv4 DNS 返回 ICP 的 IPv4 地址给用户。

⑤ 用户收到该 ICP1 的 IPv4/IPv6 地址后，优选 IPv6 地址，携带端口号发起连接。

⑥ 协议转换网关处理该连接，从 IPv4 地址池中分配一个公有 IPv4 地址 + 端口号与源 IPv6 地址 + 端口号对应，并维护一张映射表，将目的 IPv6 地址的 prefix（/96）去掉，留下 IPv4 地址作为目的地址（ICP 的 IPv4 地址）。

⑦ 协议转换网关发起与ICP1的连接，ICP1返回网页给用户，也通过协议转换网关将IPv4报头转换成 IPv6 报头。

如图 6-4 所示。

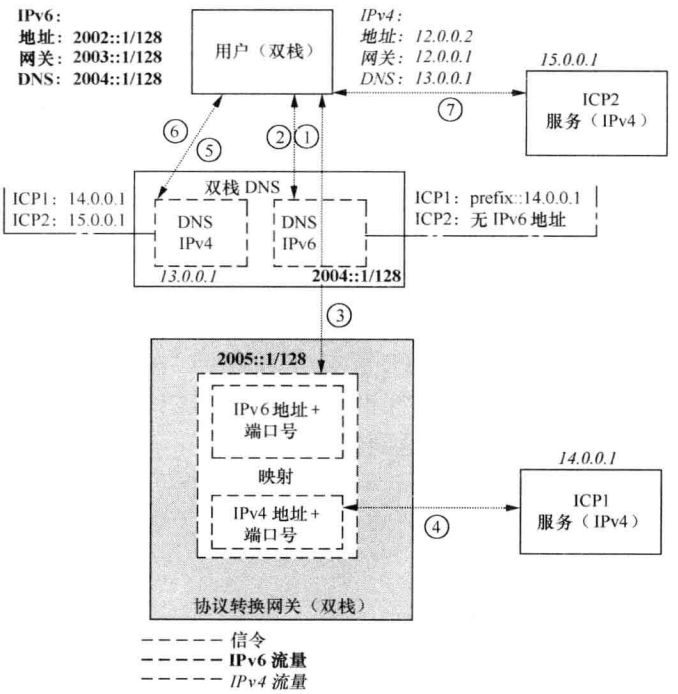


图 6-4 网页类浏览应用的协议转换方案

采用协议转换的方式只能解决使用域名访问互联网的情况，不能解决直接采用 IPv4 地址访问互联网用户的问题。

（2）运营商 Socket 代理解决方案

① 场景 1：用户是纯 IPv6 终端，直接访问一个纯 IPv4 的网站

☆ 用户使用纯 IPv6 终端上网。先进入该平台的导航网页 www.CTIPv6.com。

☆ 导航 Web 服务器通过用户源地址得知用户是纯 IPv6 用户，将 www.114.com 前加上

http://ipv6*/前缀，对 http IPv4/IPv6 协议转换服务器发送 http 请求，源地址为 IPv6 地址，并用不同的端口区分不同的 TCP 连接。

☆ http IPv4/IPv6 协议转换服务器接收到此域名后，将前缀去掉，并利用对应的 IPv4 地址向 www.114.com 发起 HTTP 请求。

☆ www.114.com 收到该请求之后返回网页首页给协议转换服务器。

☆ 服务器在该返回的 IPv4 www.114.com 网页的所有链接前加上 ipv6*/形成一个新的 IPv6 域名，其中所有的链接（包括动态链接）所对应的 IPv4 地址都改为 IPv6 地址，并推送给客户端。

☆ 客户端收到此经过协议转换的网页后，点击其中的子链接，本地操作系统发起 http 请求至协议转换网关（此时不需要经过 Portal 页面），协议网关重复第 3～5 步，将经过转换的网页发送给客户端。

② 场景 2：用户是纯 IPv6 终端，直接访问一个纯 IPv6 的网站，但其网页动态链接中包含纯 IPv4 网址

☆ 在用户第一次访问该纯 IPv6 网址时（例如：ipv6.google.com），不必经过该协议转换平台，用户可直接访问该网站。

☆ 当该网页中包含 IPv4 连接时（例如：www.114.com），此时客户端拨号软件将自动弹出导航网页。

☆ 重复场景 1 的第 3～6 步。

如图 6-5 所示。

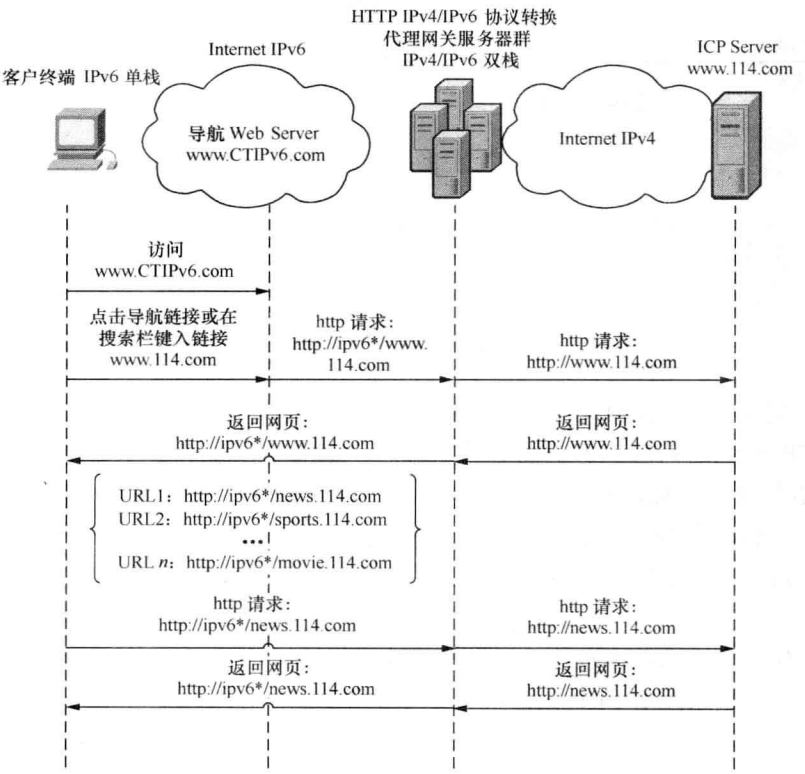


图 6-5　网页类浏览应用的 Socket 代理解决方案

使用协议转换和 Socket 代理技术只是从面向用户的角度对 ICP 进行了迁移，这两种方案只能作为过渡时期的方案，ICP 最终需要对自身的网站服务器或者程序进行改造迁移到 IPv6。

双栈的解决方案适合 ICP 对自身服务器的改造，这也是 ICP 向 IPv6 迁移必经的第一步。以下就结合 ICP 网站的架构来对不同类型的网络浏览业务的迁移来阐述对迁移改造的建议。

网页浏览业务分为新闻咨询类产品和互动类产品。新闻咨询类产品主要是静态网页，互动类产品主要是动态网页，这两类产品的双栈改造需要区别对待。

（3）ICP 新闻咨询类产品系统双栈改造解决方案

ICP 新闻咨询类产品系统 IPv4 访问流程如图 6-6 所示。

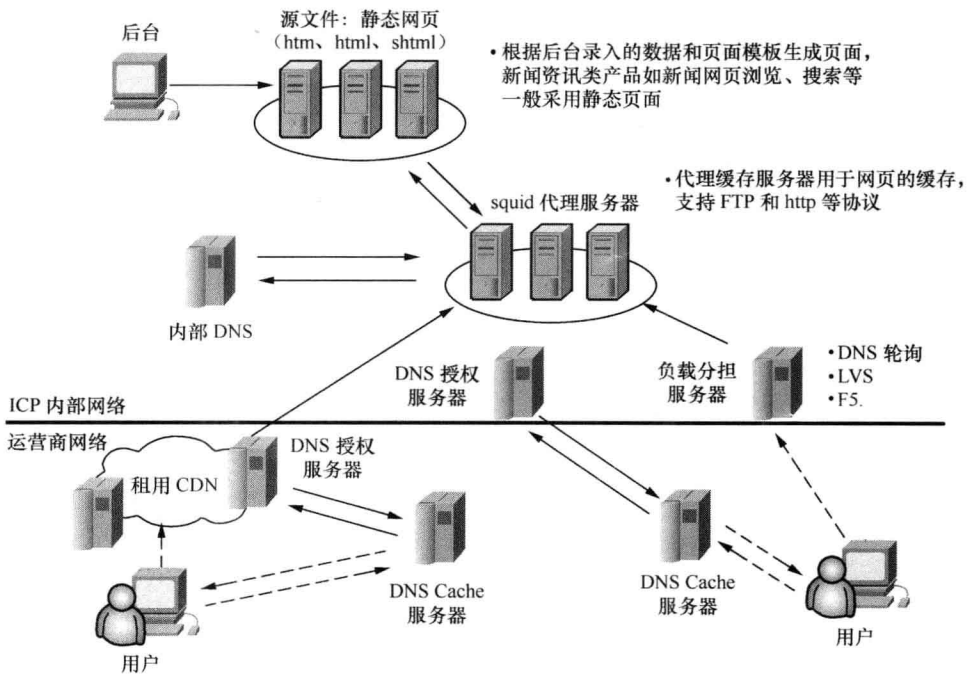


图 6-6　ICP 新闻咨询类产品系统 IPv4 访问流程图

针对新闻资讯类应用，需要将用户指向的服务器设置为双栈，即将 DNS 授权服务器中返回给用户 IP 地址所对应的服务器设置为双栈。若该服务器为负载分担服务器，则 squid 服务器也需要升级为双栈，初期可以只选择一台进行升级。若用户直接访问 ICP 的负载分担服务器，也应将此负载分担服务器设置为双栈。若用户直接访问 ICP 租用 CDN 内的服务器，则需将 CDN 内的服务器设置为双栈（需 ICP 与 CDN 服务商协商）。具体流程如图 6-7 所示。

（4）ICP 互动类产品系统双栈改造解决方案

ICP 互动类产品系统 IPv4 访问流程如图 6-8 所示。

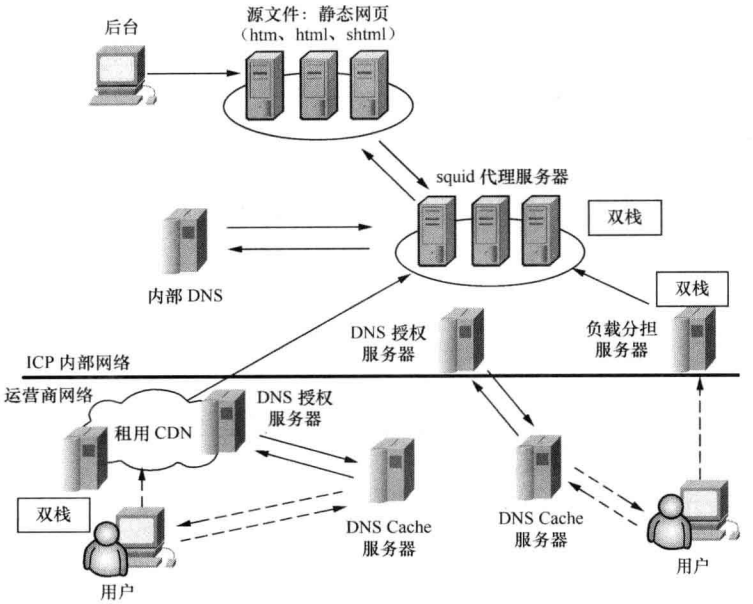


图 6-7　ICP 新闻咨询类产品系统双栈改造解决方案

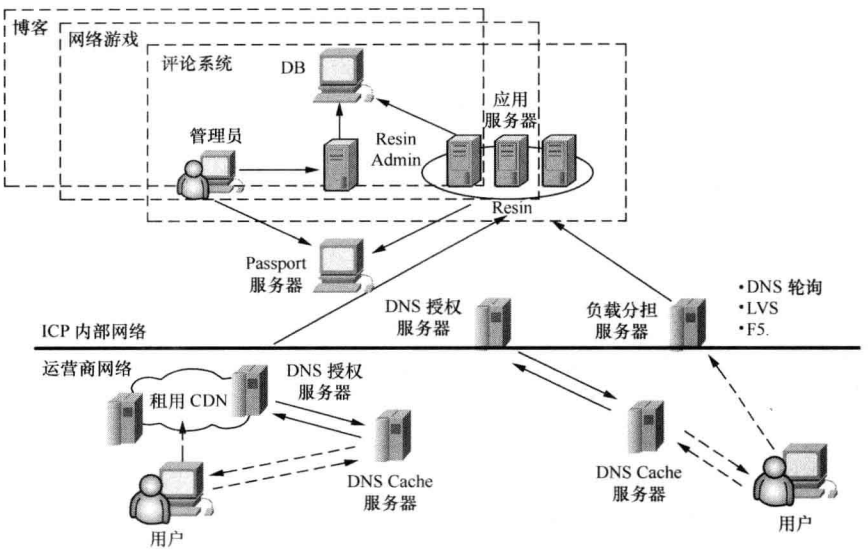


图 6-8　ICP 互动类产品系统 IPv4 访问流程图

由于互动类产品交互的信息量较少，而且不会像静态网页一样被第二次引用，因此一般用户请求不会经过 squid 服务器到达应用服务器，而会直接从负载均衡服务器发送至应用服务器。

比较典型的互动产品是论坛。在论坛中，经常需要显示用户的 IP 地址，DB（数据库）需要修改支持识别 128bit 的 IPv6 地址，同时 Apache 容器需支持 IPv6。

从网络设备方面来说，若原有网络中有 squid 服务器，squid 可将 IPv6 地址作为字符串携代在 HTTP 报头中转发出去，相当于透传 IPv6 地址，则初期可只选择一台 squid 改造。若原有网络中无 squid 服务器，则应用服务器需升级成双栈。

具体流程如图 6-9 所示。

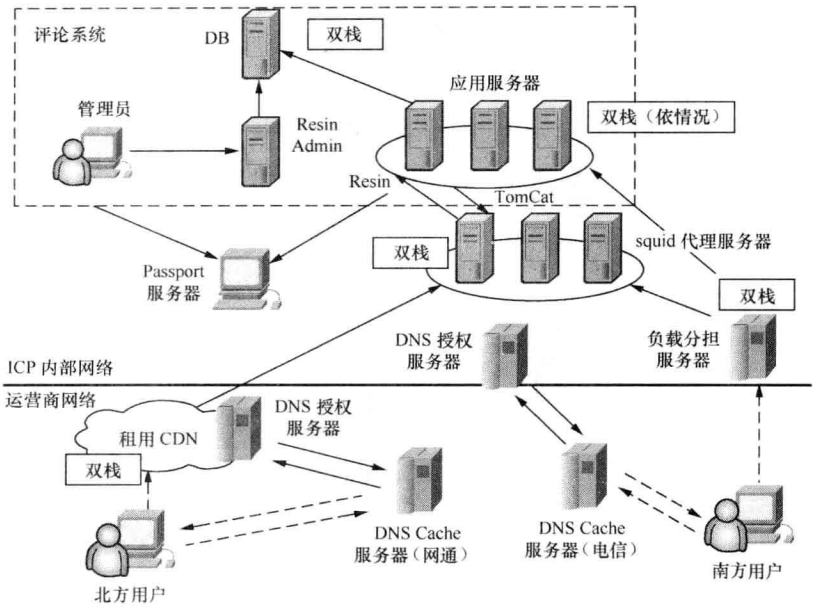


图 6-9　ICP 互动类产品系统双栈改造解决方案

6.2.2　FTP 应用迁移

1. IPv6 对 FTP 应用的影响分析

要了解 IPv6 对 FTP 的影响，首先要了解 FTP 协议的工作原理。FTP 具体工作原理如图 6-10 所示。

用户通过 FTP 传输文件存在两种形式。一种是通过域名访问服务器，另一种是直接通过 IP 地址访问服务器。通过域名访问服务器的流程见图 6-10 中实线圆圈所标识的序号，通过 IP 地址访问服务器的流程见图 6-10 中用虚线圆圈所标识的序号。

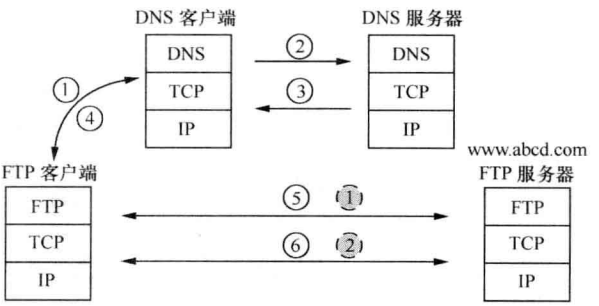


图 6-10　FTP 工作原理图

对于用户使用域名访问服务器的情况，其具体流程如下。

① 用户在 FTP 客户端中键入 www.abcd.com 之后，客户端将主机名 abcd.com 传递给本地主机上的 DNS 应用客户端。

② DNS 客户端发送 DNS 请求报文到 PC 配置好的 DNS 服务器。该请求可以是申请 IPv4 地址的 A 请求，也可以是申请 IPv6 地址的 AAAA 或者 A6 请求。

③ DNS 服务器通过域名解析，返回给 DNS 客户端该域名所对应的 IP 地址，根据申请的类型返回 IPv4 或者 IPv6 地址。

④ DNS 客户端将该 IP 地址返回给 FTP 客户端。

⑤ FTP 客户端首先与 FTP 服务器建立控制连接。它首先向服务器的 TCP 21 端口发起一个建立连接的请求，FTP 服务器接受来自客户端的请求，完成连接的建立。

⑥ FTP 控制连接建立之后，即可开始传输文件。传输文件的连接称为 FTP 数据连接。FTP 数据连接就是 FTP 传输数据的过程。它有两种传输模式：主动传输模式（Active）和被动传输模式（Passive）。在 PORT 模式（主动方式）下，FTP 客户端首先和 FTP 服务器的 TCP 21 端口建立连接，通过这个通道发送命令。客户端需要接收数据的时候在这个通道上发送 PORT 命令。PORT 命令包含了客户端用什么端口（一个大于 1024 的端口）接收数据。在传送数据的时候，服务器端通过自己的 TCP 20 端口发送数据。FTP 服务器必须和客户端建立一个新的连接用来传送数据。在被动模式下，FTP 服务器打开一个大于 1024 的随机端口，并且通知客户端在这个端口上传送数据的请求，然后 FTP 服务器将通过这个端口进行数据的传送。这个时候 FTP 服务器不再需要建立一个新的和客户端之间的连接来传送数据。

通过对用户 FTP 访问流程的分析，我们可以发现：当 IPv6 引入后，用户是否能够正常使用 FTP 进行文件传输，与 FTP 协议本身并没有关系，而是与 FTP 软件以及域名或者 IP 地址有关。同时，相关联地，也就与被访问的网站以及用户终端所支持的协议有关。

2. FTP 应用迁移方案建议

（1）运营商协议转换解决方案

同网页浏览运营商协议转换解决方案。

（2）软件改造解决方案

FTP 软件编程时将程序内嵌的 IP 地址改成域名，并在 socket 编程时指定需支持双栈。

（3）系统双栈改造解决方案

同新闻咨询类产品的系统双栈改造解决方案。

6.2.3　电子邮件应用迁移

1. IPv6 对电子邮件应用的影响分析

（1）对客户端电子邮件的影响

要了解 IPv6 对客户端邮件的影响，首先要了解客户端邮件的工作原理。客户端邮件的具体工作原理如图 6-11 所示。

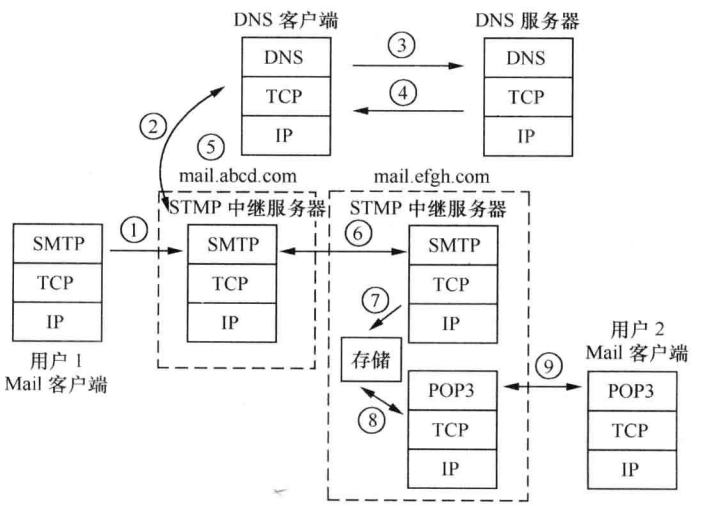


图 6-11　客户端邮件工件原理图

图 6-11 说明了用户 1 通过客户端邮件（例如 Foxmail）向用户 2 发送邮件，以及用户 2 通过客户端邮件接收该邮件的过程。这两个用户所使用的 Mail 服务器不同，分别为 mail.abcd.com 以及 mail.efgh.com。这里假设用户在邮件客户端中所定义的邮件服务器都是以域名的形式存在的。

① 用户 1 的邮件客户端程序（这里假设为 Outlook Express）与 abcd.com 的 SMTP 服务器建立网络连接，使用 SMTP 协议把邮件发送给 abcd 的 SMTP 服务器。abcd 的 SMTP 服务器收到用户 1 提交的电子邮件后，首先根据收件人的地址后缀判断接收者的邮件地址是否属于该 SMTP 服务器的管辖范围，如果是的话就直接把邮件存储到收件人的邮箱中，否则按如下流程执行。

② 图 6-11 所示第 2～5 步，abcd 的 SMTP 服务器将采用邮件传输代理（例如 Sendmail）去相应的 DNS 服务器通过 MX 记录查找用户 2 域名所对应的 SMTP 中继服务器的 IP 地址。

③ 图 6-11 所示第 6 步，abcd 与 efgh 的 SMTP 服务器建立连接并采用 SMTP 协议把邮件发送给 efgh 的 SMTP 服务器。

④ 图 6-11 所示第 7 步，efgh 的 SMTP 服务器收到 abcd 的 SMTP 服务器发来的电子邮件后，直接把邮件存储到收件人的邮箱中。

⑤ 图 6-11 所示第 8、9 步，用户 2 通过邮件客户端程序（这里假设也为 Outlook Express）与 efgh 的 POP3/IMAP 服务器建立网络连接，并以用户 2 的用户名和密码进行登录后，就可以通过 POP3 或 IMAP 协议读取邮箱中的邮件。

其中，Sendmail 的具体工作原理如下。

若 SMTP 服务器 mail.abcd.com 收到一封信要发到 user2@efgh.com。

a. Sendmail 请求 DNS 给出主机 efgh.com 的 CNAME 记录，若 CNAME 到 mail.efgh.com，则再次请求 mail.efgh.com 的 CNAME 记录，直到没有为止。

b. 假定最后 CNAME 到 mail.efgh.com，然后 Sendmail 会请求@efgh.com 域的 DNS 给出 mail.efgh.com 的 MX 记录。

c. Sendmail 最后请求 DNS 给出 mail.efgh.com 的 A/AAAA 记录，即 IPv4 或者 IPv6 地址。

d. Sendmail 与该 IP 地址连接，传送这封给 user2@efgh.com 的信到该 IP 地址服务器的 SMTP 后台程序。

通过对客户端邮件收发流程的分析，我们可以发现：当 IPv6 引入后，用户是否能够成功使用客户端，与 SMTP 协议以及 POP3 协议本身并没有关系，而是与客户端软件、邮件服务器是以域名还是 IP 地址的方式呈现，以及 SMTP 服务器的 Sendmail 等邮件传输代理软件有关。

如果用户的邮件客户端所配置的邮件服务器地址是 IPv4 地址，则 IPv6 单栈 PC 无法成功访问邮件服务器。

如果客户端软件以及 Sendmail 等邮件转发软件在 Socket 编程时，指定了 IP 层协议选择 IPv4（VC＋＋编程）或者指定了 DNS 返回的地址只能是 IPv4 地址（Java 编程），则即便客户端软件是支持 IPv6 的，用户也不能正常的发送邮件。

（2）对 Web 电子邮件的影响

用户采用 http 或者 https 的方式来访问 Web 邮箱，从浏览器到邮件服务器的访问过程与网页浏览一致，从 A 网站服务器到 B 网站服务器的转发过程与客户端邮件一致。因此可以知

道，IPv6 对 Web 邮件的影响，只与浏览器软件与邮件转发软件对底层 IP 协议的调用有关。

2．电子邮件应用迁移方案建议

图 6-12 是一个典型的邮件服务系统架构图。面向用户的邮件服务器地址可以是邮件服务 ICP 本身的负载分担服务器的地址，也有可能是该 ICP 在 CDN 所租用的地址。用户在发送或者接收邮件时，配置的邮件服务器地址一般是面向用户的负载分担服务器。负载分担服务器在接收到用户的发送、接收邮件或者邮件网页访问的请求后，将不同类型的报文分发到 SMTP、POP3 以及 Web mail 服务器进行处理。各服务器再到存储服务器群上传或者下载所需的邮件。

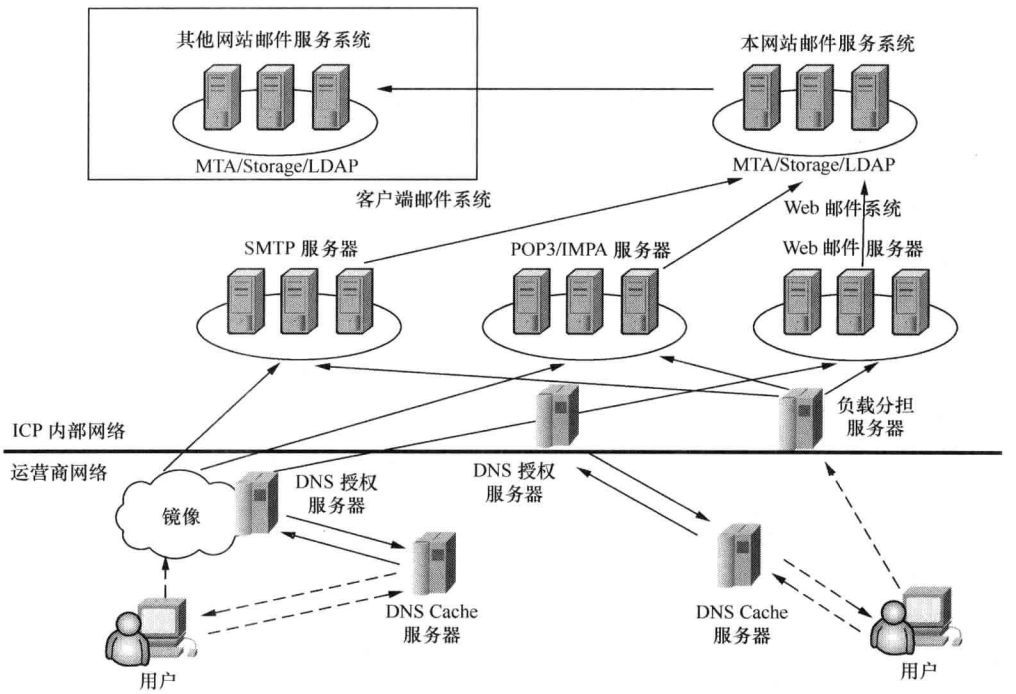


图 6-12　邮件服务系统架构图

（1）客户端电子邮件迁移方案

① 运营商协议转换解决方案

同网页浏览运营商协议转换解决方案。

② 软件改造解决方案

客户端软件以及 Sendmail 等邮件转发软件在 Socket 编程时，应指定 IP 层协议有 IPv4 和 IPv6 两个接口（VC + +编程），或者指定 DNS 返回的地址可以返回 IPv4 或者 IPv6 地址（Java 编程）。

③ 系统改造解决方案

将用户指向的服务器设置为双栈即可，即将 DNS 授权服务器中返回给用户 IP 地址所对应的服务器设置为双栈。本例中将负载分担服务器和北方镜像服务器设置成双栈。

对于客户端收发邮件的情况，需要将 SMTP 和 POP3 服务器升级为双栈或者新建。

具体改造情况如图 6-13 所示。

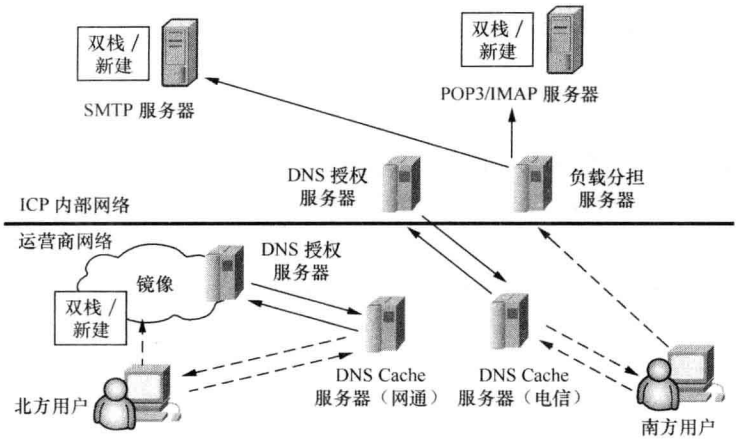


图 6-13　客户端电子邮件迁移方案

（2）Web 邮件迁移方案

① 运营商协议转换解决方案

同网页浏览运营商协议转换解决方案。

② 运营商 Socket 代理解决方案

同网页浏览运营商 Socket 代理解决方案。

③ 软件改造解决方案

Sendmail 等邮件转发软件在 Socket 编程时，应指定 IP 层协议有 IPv4 和 IPv6 两个接口（VC＋＋编程），或者指定 DNS 返回的地址可以返回 IPv4 或者 IPv6 地址（Java 编程）。

④ 系统改造解决方案

要将用户指向的服务器设置为双栈，即将 DNS 授权服务器中返回给用户 IP 地址所对应的服务器设置为双栈。本例中将负载分担服务器和北方镜像服务器设置成双栈。

对于 Web mail 服务器，可通过加入 squid 代理服务器访问。squid 代理服务器升级为双栈，Web mail 服务器无需升级；若不采用 squid 服务器，则全部 Web mail 服务器都需升级。

具体情况如图 6-14 所示。

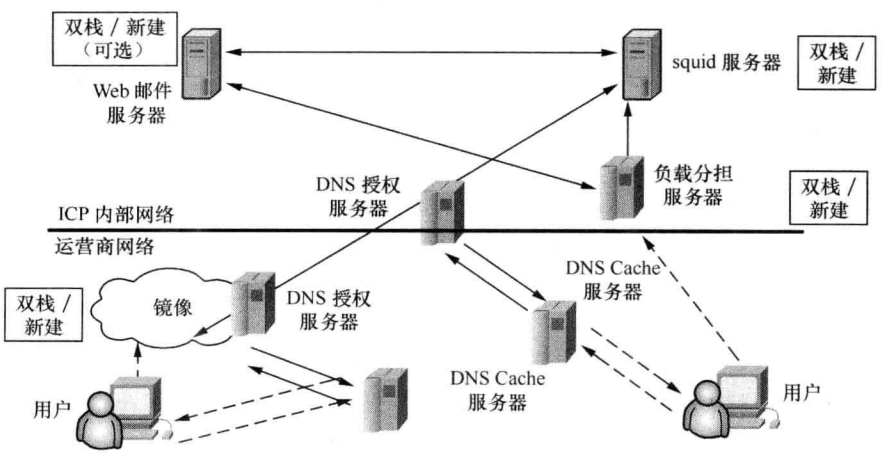


图 6-14　Web 邮件迁移方案

6.2.4　即时通信应用迁移

1．IPv6 对即时通信应用的影响分析

要了解 IPv6 对即时通信的影响，首先要了解即时通信的工作原理。现有的即时通信协议大部分都是私有协议，但总体流程基本一致，如图 6-15 所示。

首先是用户登录。当用户登录时，即时通信软件会向其内设的登录服务器发起请求。若该登录服务器不是使用域名而是使用 IP 地址，则当用户是 IPv6 单栈时，会存在服务器无法连接的问题。若登录成功，用户之间的文本交互还需要通过另外的接线服务器。该服务器的地址也是内设在即时通信软件中的。可以看到，如果即时通信软件内容的服务器都是以 IP 地址的形式存在的话，当用户终端是 IPv6 单栈时，都会出现服务器连接不上的问题。

具体的即时通信工具登录过程如图 6-16 所示。

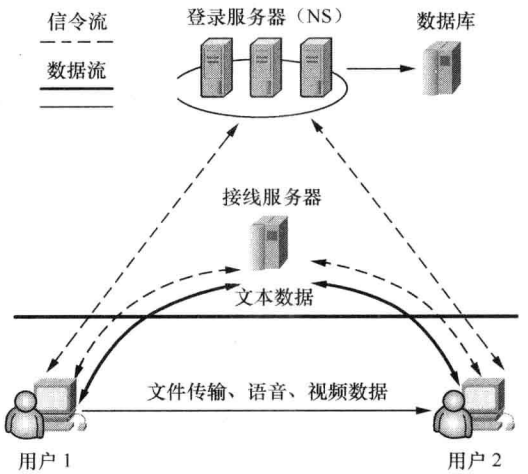


图 6-15　即时通信工作原理图

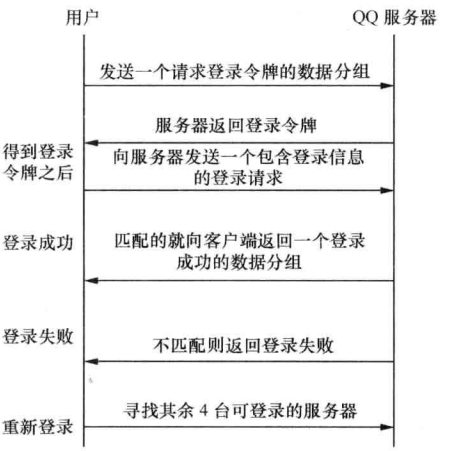


图 6-16　即时通信工具用户登录过程

具体情况如下。

① 首先客户端向服务器发送一个请求登录令牌的数据分组。服务器返回登录令牌。这个令牌是在服务器端生成的，和客户端的 IP 地址、版本信息等数据相关。以后会用到这个令牌去进行其他操作。

② 在客户端得到登录令牌之后，就会向服务器发送一个包含登录信息的登录请求，要求登录。服务器首先会确认客户端的号码、地址和版本是否可以在本服务器上进行登录。

③ 如果可以的话，就验证客户端的登录信息是否与服务器上保存的登录信息相匹配。匹配的就向客户端返回一个登录成功的数据分组。不匹配则返回登录失败。

④ 因为服务器有很多台，可能要分管不同的即时通信工具版本、IP 地址等信息，所以如果客户端的号码、IP 地址和版本无法在本服务器进行登录，服务器就返回一个重定向分组，让客户端去另外一台服务器登录。

从用户登录的过程可以看出，引入 IPv6，除了软件内嵌的服务器采用 IP 地址登录的方式会导致服务器连接失败之外，由于用户地址变成了 IPv6 地址，与用户 IP 地址相关联的令牌如果只能进行 IPv4 地址关联，也会导致认证失败。

另外，对于文件传输以及音频、视频服务，其流程与文本交互对比关系如图 6-17 所示。

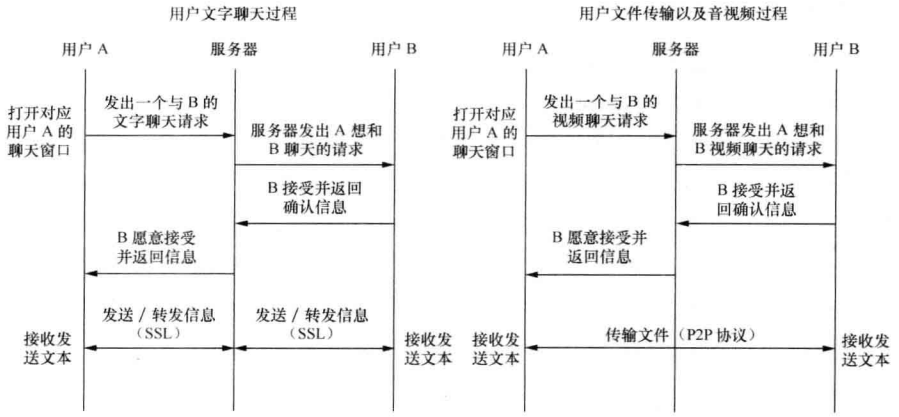


图 6-17　文件传输及音视频服务的流程与文本交互对比关系图

当用户登录成功后，开展的业务有文字聊天以及文件传输等。文字聊天全部是 C-S 模式通过服务器中转。文件传输以及音频、视频服务在信令建立过程中是使用 C-S 模式由服务器进行代理的，但是数据传输是采用点到点的模式。

引入 IPv6 后，除了会因服务器使用 IPv4 地址而使用 IPv4 地址出现服务器连接失败的情况外，最关键的是点到点的文件和音频、视频传输会出现问题。当两端的用户不是使用相同的 IP 协议时，传输会失败。

2．即时通信应用迁移方案建议

（1）软件改造迁移方案

① 令牌生成程序

用户端登录时，服务器返回给用户登录令牌。这个令牌是在服务器端生成的，和客户端的 IP 地址、版本信息等数据相关。因此程序需要改造识别 128bit 的 IPv6 地址。

② DB 数据库程序

服务器会根据客户端的号码、IP 地址和版本查询数据库，以确认该用户是否可以在本服务器上进行登录。因此数据库需要改造识别 IPv6 地址，又分成两种方案。

☆ 方案一：用户采用 IPv4/IPv6 同时登录。

a．中转服务器程序：当 IPv6 用户请求与 IPv4 用户建立连接的时候，中转服务器需要根据 DB 数据库中的用户数据，返回 IPv4 用户不可达的信息给 IPv6 客户端。

b．客户端：收到 IPv4 用户不可达信息之后，可改造使用 IPv4 协议发起连接。

☆ 方案二：用户采用 IPv4 或 IPv6 单栈登录。

a．中转服务器程序：当用户请求与不同质用户建立连接的时候，中转服务器需要根据 DB 数据库中的用户数据，返回用户不可达的信息给 IPv6 客户端。

b．客户端：若 ICP 建设有协议转换服务器，则 IPv6 客户端通过该协议转换服务器向 IPv4 用户发起新的连接，IPv4 客户端无需改造。若网络中没有协议转换服务器，则客户端无需改造。

（2）系统改造迁移方案

① 各类型服务器皆需升级至 IPv6 或者新建。各类型服务器群中初期可以选择一台进行升级。

② 文本和音频、视频传输都是 P2P 模型，因此初期可先不考虑迁移。

6.2.5　流媒体应用迁移

1．IPv6 对流媒体应用的影响分析

在流媒体传输中，标准的协议是 RTP（Real time Transport Protocol，实时传输协议）、RTCP（Real-time Transport Control Protocol，实时传输控制协议）、RTSP（Real Time Streaming Protocol，实时流媒体协议）和 RSVP（Resource Reserve Protocol，资源预订协议）。

经过厂商的基于标准协议的开发，现在主流的流媒体传输协议有以下几种。

① RTSP（Real Time Streaming Protocol），实时流媒体协议，它是由 RealNetworks 和 Netscape 共同提出的，现在用于 RealNetworks 的 Real Media 产品中；

② PNM（Progressive Networks Audio），这也是 Real 专用的实时传输协议。它一般采用 UDP 协议，并占用 7070 端口。但当你的服务器在防火墙内且 7070 端口被挡，且你的服务器把 SmartingNetwork 设为真时，则采用 http 协议，并占用默认的 80 端口。

③ MMS（Microsoft Media Server protocol），这是微软的流媒体服务器协议。MMS 是连接 Windows Media 单播服务的默认方法。微软没有公开此协议，属于私有协议。

用户主要有两种使用流媒体技术的方式。一种是通过浏览器观看在线 Web 视频，另外一种是通过本地客户端的方式观看视频。现在视频主要采用 RTSP 等流媒体技术，而客户端模式主要采用 P2P 技术。如图 6-18 所示。

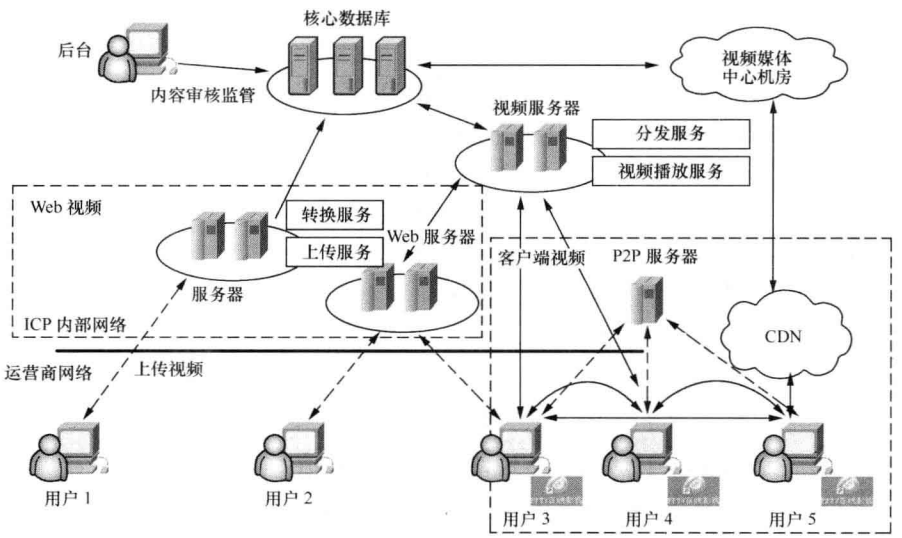


图 6-18　流媒体应用

（1）对 Web 流媒体应用的影响

要了解 IPv6 对 Web 视频的影响，首先要了解 Web 视频的工作原理。Web 视频一般采用内嵌 RTSP（Real Time Streaming Protocol，实时流媒体协议）技术和 RTP 技术的浏览器来实现在线观看。RTSP 属于业务控制协议，基于 TCP，它定义了如何在 IP 网络上有效地传输流媒体数据。RTSP 提供了一种机制，使音频、视频等数据可以按照需要进行实时传输，并且可以实施诸如暂停、快进等控制。源数据可以是存储的文件，也可以是现场数据的反馈。而数据传输本身是由基于 UDP 的 RTP 协议来完成的。

RTSP 协议与 HTTP 协议有点类似，通信双方是通过请求消息和回应消息进行交互的。不同的是 http 是由客户端向服务器发起的单向请求，但是 RTSP 可以由客户端和服务器双向发起请求。一个典型的 RTSP 的交互过程如图 6-19 所示，其中，RTSP 客户端指的是嵌入了 RTSP 模块的浏览器。

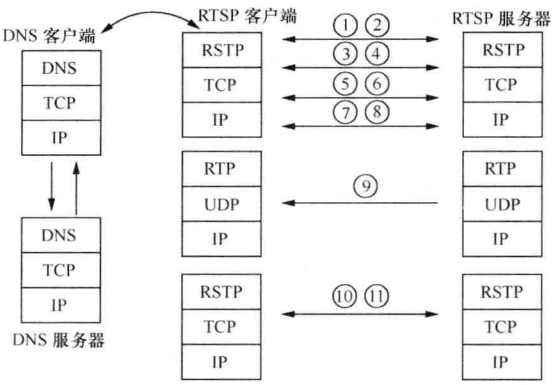


图 6-19　典型的 RTSP 协议交互过程

具体流程如下。

① RTSP 客户端向 RTSP 服务器发送 OPTION 请求，询问服务器有哪些方法可以用。

② RTSP 服务器回应客户端 OPTION 请求，包含了所有可用的方法，比如 DESCRIBE、PLAY、PAUSE 等。

③ RTSP 客户端向 RTSP 服务器发送 DESCRIBE 请求，该请求中携带了视频源目的域名或者地址，要求服务器提供媒体初始化的描述信息。

④ 服务器回应客户端的媒体初始化信息，说明服务器端的各个参数。

⑤ 客户端向服务器发起 SETUP 请求，请求建立连接，设置会话属性以及传输模式。

⑥ 服务器向客户端发起 SETUP 回应，建立会话，返回会话标识符，以及会话相关信息。

⑦ 客户端向服务器发起 PLAY 请求，请求建立播放。

⑧ 服务器向客户端发起 PLAY 回应。

⑨ 通过 RTP 传输数据流。

⑩ 客户端向服务器发起 TEARDOWN 请求，请求关闭会话。

⑪ 服务器向客户端发起 TEARDOWN 回应，关闭会话。

步骤③中，在 RTSP 客户端向 RTSP 服务器发送 DESCRIBE 请求时，具体信息如下。

C→S: DESCRIBE rtsp://server.example.com/fizzle/foo RTSP/1.0

CSeq: 312

Accept: application/sdp, application/rtsl, application/mheg

这时 RTSP 客户端会调用本机的 DNS 客户端来进行域名解析，得到目的 IP 地址。这里 rtsp://后面也可以直接填入 IP 地址，在这种情况下，对于纯 IPv6 客户端，一个 IPv4 地址会导致连接失败。

步骤④中，RTSP 服务器回应 RTSP 客户端的媒体初始化信息时，具体情况如下。

S→C:　　RTSP/1.0 200 OK

CSeq: 312

Date: 23 Jan 1997 15:35:06 GMT
Content-Type: application/sdp
Content-Length: 376
v = 0
o = mhandley 2890844526 2890842807 IN IP4 126.16.64.4
s = SDP Seminar
i = A Seminar on the session description protocol
u = http://www.cs.ucl.ac.uk/staff/M.Handley/sdp.03.ps
e = mjh@isi.edu（Mark Handley）
c = IN IP4 224.2.17.12/127
t = 2873397496 2873404696
a = recvonly
m = audio 3456 RTP/AVP 0
m = video 2232 RTP/AVP 31
m = whiteboard 32416 UDP WB
a = orient:portrait

在使用 RTSP 协议进行交互的过程中，存在报文中携带 IP 地址的情况，对于 IP 地址不同质的终端，会存在访问失败的问题。

（2）对客户端流媒体应用的影响

流媒体客户端除了类似暴风影音以及 PPLive 之类的客户端视频外，还包括内嵌在 IE 中的视频加速器。客户端视频有两种通信模式，一种是 C/S 模式，还有一种更加常用的是 P2P 模式。P2P 应用与 IP 地址相关性更强，对纯 IPv6 用户正常访问的影响更大。

2．流媒体应用迁移方案建议

（1）Web 流媒体应用迁移方案建议

Web 流媒体应用指的是在用户没有安装任何插件的情况下通过浏览器访问的视频。插件归属与客户端视频软件。

对于上传用户和下载用户，要将用户指向的服务器升级为双栈或者新建，即将 DNS 授权服务器中返回给用户 IP 地址所对应的服务器设置为双栈或者新建。设备可以是负载分担服务器，可能是 Squid 服务器，也有可能就是 Web 服务器（视各网站具体情况而定）。

因此，在 Web 流媒体应用环境下，相关应用需做以下调整。

① ICP 租用或者自建的 CDN 的服务器需升级为双栈或者新建。

② 采用的协议为 http，无需修改协议。

具体情况如图 6-20 所示。

（2）客户端流媒体应用迁移方案建议

客户端视频软件除了通常意义上的客户端软件（如 PPLive）之外，还包括 Web 视频加速插件。通过客户端软件观看视频的时候，包括 C/S（mms、RTP 和 RTCP）和 P2P 两种模式。

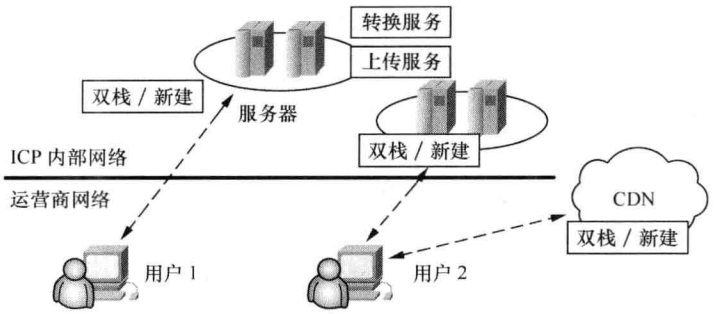


图 6-20　Web 流媒体应用迁移方案

① C/S 模式：除了客户端支持双栈之外，要求用户所指向的服务器支持双栈或者新建（在 mms、RTP 以及 RTSP 等标准协议支持 IPv6 的情况下，若采用私有协议，则需要修改私有协议支持 IPv6）。

② P2P 模式：包括加速插件和 P2P、P4P 浏览器（如 PPLive）。除了客户端支持双栈之外，要求 P2P 种子服务器支持双栈或新建，并且可以根据用户 IP 地址类型，返回相同类型的种子地址（比如给 IPv6 用户返回 IPv6 的种子地址）。建议客户端软件改造为具备同时发起 IPv4 和 IPv6 P2P 链接的功能，并可将接收回来的视频块共享给 IPv4 和 IPv6 两个 P2P 链接（如某视频块只有 IPv4 种子而没有 IPv6 的种子，用户通过 IPv4 下载这个视频块；上传时，这个视频块通过 IPv4/IPv6 链接同时上传，这样通过客户端改造可自动地将 IPv4 视频资源分发到 IPv6 平面），以规避 IPv6 初期种子数过少而导致的视频几乎无法下载的情况。

6.3　IPv6 与物联网应用

6.3.1　物联网概念与架构

1．物联网基本概念

物联网（The Internet of things），通俗来讲，就是“物物相联的互联网”。严格而言，物联网的定义在于：通过信息传感设备（如射频识别（RFID）、红外感应器、全球定位系统、激光扫描器等），按照约定的协议，把物品与互联网连接起来，进行信息交换和通信，以实现智能化识别、定位、跟踪和管理的一种网络。

也许，单纯的概念无法让你理解物联网到底是什么。那么可以通过生活中的例子来了解物联网。大家可能都能理解互联网，互联网即是 IT 设施通过网络互联。而我们可以将物联网认为是一种特殊的互联网，在这种互联网中，网络互联的不仅仅包括 IT 设施，更包括普通的物理设施，例如建筑物、杯子、桌子。当互联网的终端设备扩大到普通的物理设施时，自然地产生了物联网。

2．物联网的实质

物联网是在计算机互联网的基础上，利用 RFID、无线通信等技术，构造一个覆盖到万物的互联网络。在这样的网络中，普通的物品能够彼此交流，无须人的干预。其实质在于利用射频自动识别技术，通过话机互联网实现物品的自动识别和信息的互联共享。

而射频自动识别（RFID）技术，正是能够让物品“开口说话”的一种技术。在物联网这

一构想中，RFID 标签中存储着规范而具有互用性的信息，通过无线通信网络把它们自动采集到中央信息系统，实现物品的识别，进而通过开放性的话机网络实现信息交换和共享。

物联网的概念打破了传统思维，将物理基础设施和 IT 基础设施整合起来，使得钢筋混凝土与电脑芯片成为统一的基础设施，在物联网的世界里平等对话。

3．物联网体系架构

物联网可分为 3 层：感知层、网络层和应用层。通俗地说，感知层相当于人体的皮肤和五官、网络层相当于人的神经中枢和大脑，而应用层相当于人的社会分工。

① 感知层：识别物体，采集信息。感知层一般包括二维码标签和识读器、RFID 标签和读写器、摄像头、GPS、传感器、终端、传感器网络等，其作用与人体结构中皮肤和五官的作用相似。

② 网络层：信息传递和处理。网络层包括通信与互联网的融合网络、网络管理中心、信息中心和智能处理中心等。网络层将感知层获取的信息进行传递和处理，类似于人体结构中的神经中枢和大脑。

③ 应用层：与行业需求结合，实现广泛智能化，是对网络层传递和处理的信息进行应用，如实现工业监控、环境监测等。

物联网的三层体系架构如图 6-21 所示。

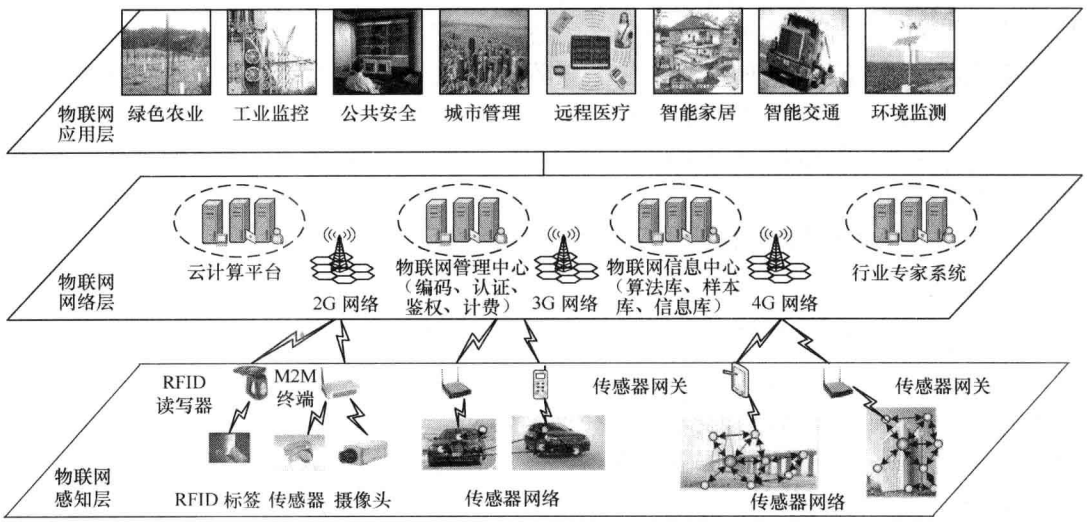


图 6-21　物联网的体系架构

4．认识 M2M

M2M 是机器对机器（Machine-to-Machine）通信的简称。它是一种理念，也是所有增强机器设备的网络通信能力的技术的总称。M2M 技术的目标就是使所有的机器设备都具备联网的通信能力，其核心理念就是 Network Everything。

简单来说，M2M 就是机器的互联网。实现手段包括各种联网的技术，实现形式包括人对机器、机器对人，以及机器对机器的通信。而物联网是比 M2M 更加广泛的概念，M2M 是实现物联网的第一步。严格来讲，M2M 作为一种技术的总称，是实现物联网的方式。

任何一种 M2M 技术与应用，都涉及 5 个重要的技术部分：机器、M2M 硬件、通信网络、

中间件、应用，如图 6-22 所示。

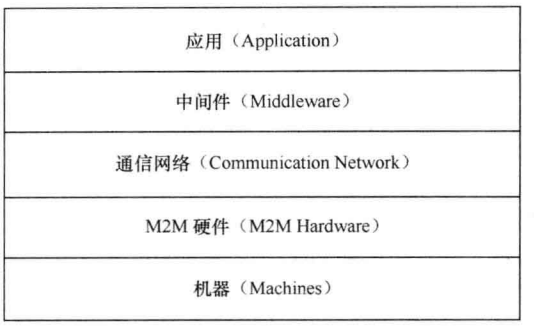


图 6-22　M2M 系统组成部分

① 机器：实现 M2M 的第一步就是从机器/设备中获得数据，然后把它们通过网络发送出去。使机器具备“说话”（talk）能力的基本方法有两种：生产设备的时候嵌入 M2M 硬件；对已有机器进行改装，使其具备通信和联网能力。

② M2M 硬件：M2M 硬件是使机器获得远程通信和联网能力的部件。现在的 M2M 硬件产品可分为 5 种，即，嵌入式硬件、可组装硬件、调制解调器、传感器、RFID。这些 M2M 硬件的共性是进行信息的提取，从各种机器/设备那里获取数据，并传送到通信网络。

③ 通信网络：通信网络在整个 M2M 技术框架中处于核心地位，包括广域网和局域网。在 M2M 技术框架的通信网络中，有两个主要参与者，他们是网络运营商和网络集成商。尤其是移动通信网络运营商，在推动 M2M 技术应用方面起着至关重要的作用，他们是 M2M 技术应用的主要推动者。第三代移动通信技术除了提供语音服务之外，数据服务业务的开拓是其发展的重点。随着移动通信技术向 3G 的演进，必定将 M2M 应用带到一个新的境界。

④ 中间件：中间件包括两部分，即，M2M 网关、数据收集/集成部件。网关是 M2M 系统中的“翻译员”，它获取来自通信网络的数据，将数据传送给信息处理系统，其主要的功能是完成不同通信协议之间的转换。数据收集/集成部件是为了将数据变成有价值的信息，对原始数据进行不同方式的加工和处理，并将结果呈现给需要这些信息的观察者和决策者。这些中间件包括数据分析和商业智能部件、异常情况报告和工作流程部件、数据仓库和存储部件等。

⑤ 应用：将获取到的数据用于各种需要的行业和用途。

5. M2M 应用发展

M2M 技术的应用很多，包括自动抄表、标志和照明管理、设备监测、环境监测、智能家居、安全防护、智能建筑、移动 POS 机、移动售货机、车队管理等，而且每一个应用又可以拓展为几个子应用，比如环境监测又可分为水质监测、温度监测等。这些应用可以应用在能源、政府、医疗、工业等行业。虽然业务很多，但大多是面向政府和企业的，面向大众的较少。M2M 主要应用如图 6-23 所示。

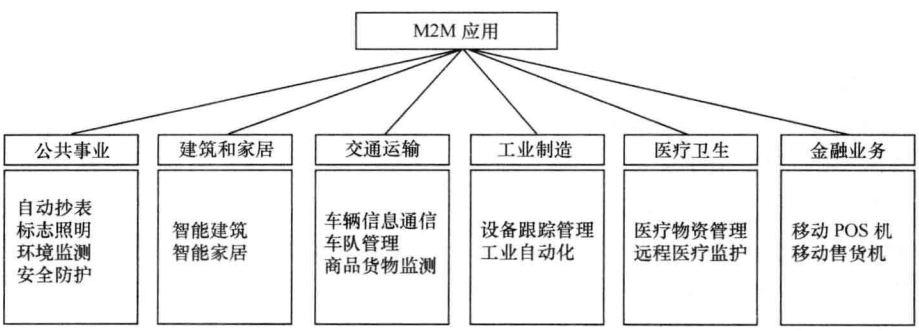


图 6-23　M2M 应用分类

6.3.2　物联业务驱动的 IPv6 演进

1．IPv6 对物联网的支持

物联网在不断的发展过程中必然会产生一些突出的需求。其一，众多的物体接入互联网并实现对它们的寻址需要巨大的地址空间；其二，要实现众多物与物、物与人之间的有效通信，必须建立统一的通信平台；其三，物联网对通信安全及质量保证必然会有较高的要求。

互联网 TCP/IP 协议族正是一种统一的互联网通信协议标准。IPv6 的引入将使大量的、多样化的终端更容易接入 IP 网络，并极大地增强了网络安全性和终端移动性。基于 IPv6 的物联网，可以在 IP 层上对数据分组进行高强度的安全处理，使用 AH 报头、ESP 报头来保护 IP 通信的安全，其安全机制更加完善（相对于 IPv4 来说）。同时，终端移动性更有利于监测物品的实时位置。IPv6 将促进物联网向着更便捷、更安全的方向发展。

（1）IPv6 在物联网寻址中的优势

物联网由众多的节点连接构成，无论是采用自组织方式，还是采用现有的公众网进行连接，这些节点之间的通信必然涉及寻址问题。美国权威咨询机构 Forrester 预测，2020 年，世界上物物互联的业务，和人与人通信的业务相比，将达到 30∶1，仅仅是在智能电网和机场防入侵系统方面的市场就有上千亿元。对于如此大的市场需求，我们不难预测物联网将要承载的对象数量之庞大，而要实现如此庞大数量的对象之间的有效通信，寻址绝不是一个简单的问题。

随着互联网本身的快速发展，IPv4 的地址已经日渐匮乏。从目前的地址消耗速度来看，IPv4 地址空间已经很难再满足物联网对网络地址的庞大需求。从另一方面来看，物联网对海量地址的需求，也对地址分配方式提出了要求。海量地址的分配无法使用手工分配，使用传统 DHCP 的分配方式对网络中的 DHCP 服务器也提出了极高的性能和可靠性要求，可能造成 DHCP 服务器性能不足，成为网络应用的一个瓶颈。

IPv6 拥有巨大的地址空间，同时 128bit 的 IPv6 的地址被划分成两部分，即地址前缀和接口地址。与 IPv4 地址划分不同的是，IPv6 地址的划分严格按照地址的位数来进行，而不采用 IPv4 中的子网掩码来区分网络号和主机号。IPv6 地址的前 64bit 被定义为地址前缀。地址前缀用来表示该地址所属的子网络，即地址前缀用来在整个 IPv6 网中进行路由。而地址的后 64bit 被定义为接口地址，接口地址用来在子网络中标识节点。在物联网应用中可以使用 IPv6 地址中的接口地址来标识节点。在同一子网络下，可以标识 264 个节点。这个标识空间约有 185 亿亿个地址空间。这样的地址空间完全可以满足节点标识的需要。

IPv6 采用了无状态地址分配的方案来解决高效率海量地址分配的问题。其基本思想是网络侧不管理 IPv6 地址的状态，包括节点应该使用什么样的地址、地址的有效期有多长，且基本不参与地址的分配过程。节点设备连接到网络中后，将自动选择接口地址（通过算法生成 IPv6 地址的后 64bit），并加上 FE80 的前缀地址，作为节点的本地链路地址。本地链路地址只在节点与邻居之间的通信中有效，路由器设备将不路由以该地址为源地址的数据分组。在生成本地链路地址后，节点将进行 DAD（地址冲突检测），检测该接口地址是否有邻居节点已经使用。如果节点发现地址冲突，则无状态地址分配过程将终止，节点将等待手工配置 IPv6 地址。如果在检测定时器超时后仍没有发现地址冲突，则节点认为该地址可以使

用，此时终端将发送路由器前缀通告请求，寻找网络中的路由设备。当网络中配置的路由设备接收到该请求时，则将发送地址前缀通告响应，将节点应该配置的 IPv6 地址前 64bit 的地址前缀通告给网络节点。网络节点将地址前缀与接口地址组合，构成节点自身的全球 IPv6 地址。

采用无状态地址分配之后，网络侧不再需要保存节点的地址状态，维护地址的更新周期。这大大简化了地址分配的过程。网络可以以很低的资源消耗来达到海量地址分配的目的。

（2）IPv6 对物联网节点移动性的支持

根据物联网的定义可知，物联网所要实现的物与物之间的通信基本上是基于无线传感技术的。也就是说物联网相对于传统的互联网对移动通信性能有了更高的要求，可以说物联网是一个瞬息万变的网络。而事实上，将来主宰物联网世界的必定是如今的移动通信服务供应商。

目前互联网的移动性不足造成了物联网移动能力的瓶颈。IPv4 协议在设计之初并没有充分考虑到节点移动性带来的路由问题，即当一个节点离开了它原有的网络，如何再保证这个节点访问可达性的问题。由于 IP 网络路由的聚合特性，在网络路由器中路由条目都是按子网来进行汇聚的。

当节点离开原有网络，其原来的 IP 地址离开了该子网，而节点移动到目的子网后，网络路由器设备的路由表中并没有该节点的路由信息（为了不破坏全网路由的汇聚，也不允许目的子网中存在移动节点的路由），会导致外部节点无法找到移动后的节点。因此，支持节点移动的能力是需要通过特殊机制实现的。在 IPv4 中 IETF 提出了 MIPv4（移动 IP）的机制来支持节点的移动。但这样的机制引入了著名的三角路由问题。对于少量节点的移动，该问题引起的网络资源损耗较小。而对于大量节点的移动，特别是物联网中特有的节点群移动和层移动，会导致网络资源被迅速耗尽，使网络处于瘫痪的状态。

IPv6 协议设计之初就充分考虑了对移动性的支持。针对移动 IPv4 网络中的三角路由问题，移动 IPv6 提出了相应的解决方案。

首先，从终端角度，IPv6 提出了 IP 地址绑定缓冲的概念，即 IPv6 协议栈在转发数据分组之前需要查询 IPv6 数据分组目的地址的绑定地址。如果查询到绑定缓冲中目的 IPv6 地址存在绑定的转交地址，则直接使用这个转交地址为数据分组的目的地址。这样发送的数据流量就不会再经过移动节点的家乡代理，而直接转发到移动节点本身。

其次，MIPv6 引入了探测节点移动的特殊方法，即某一区域的接入路由器以一定时间进行路由器接口的前缀地址通告。当移动节点发现路由器前缀通告发生变化，则表明节点已经移动到新的接入区域。与此同时根据移动节点获得的通告，节点又可以生成新的转交地址，并将其注册到家乡代理上。

MIPv6 的数据流量可以直接发送到移动节点，而 MIPv4 的数据流量必须经过家乡代理的转发。在物联网应用中，传感器有可能密集地部署在一个移动物体上。例如为了监控地铁的运行参数等，需要在地铁车厢内部署许多传感器。从整体上来看，地铁的移动就等同于一群传感器的移动。在移动过程中必然发生传感器的群体切换。在 MIPv4 的情况下，每个传感器都需要建立到家乡代理的隧道连接，这样对网络资源的消耗非常大，很容易导致网络资源耗尽而瘫痪。在 MIPv6 的网络中，传感器进行群切换时只需要向家乡代理注册，之后的通信完全在传感器和数据采集的设备之间直接进行。这样就可以使网络资源消耗的压力大大下降。

因此，在大规模部署物联网应用，特别是移动物联网应用时，MIPv6 是一项关键性的技术。

（3）IPv6 在保证物联网网络质量中的优势

网络质量保证也是物联网发展过程中必须解决的问题。目前 IPv4 网络中实现 QoS 有两种技术。其一，采用资源预留（Interserv）的方式，利用 RSVP 等协议为数据流保留一定的网络资源。在数据分组传送过程中保证其传输的质量；其二，采用 Diffserv 技术，由 IP 分组自身携带优先级标记，网络设备根据这些优先级标记来决定分组的转发优先策略。目前 IPv4 网络中服务质量的划分基本是从流的类型出发。使用 Diffserv 来实现端到端的服务质量保证。例如，视频业务有低分组丢失、时延、抖动的要求，就给它分配较高的服务质量等级；数据业务对分组丢失、时延、抖动不敏感，就分配较低的服务质量等级。但这样的分配方式仅考虑了业务的网络侧质量需求，没有考虑业务应用侧的质量需求。例如，一个普通视频业务对服务质量的需求可能比一个基于物联网传感的手术应用对服务质量的需求要低。因此，物联网中的服务质量保障必须与具体的应用相结合。

在网络服务质量保障方面，IPv6 在其数据分组结构中定义了流量类别字段和流标签字段。流量类别字段有 8bit，和 IPv4 的服务类型（ToS）字段功能相同，用于对报文的业务类别进行标识；流标签字段有 20bit，用于标识属于同一业务流的分组。流标签和源、目的地址一起，唯一标识了一个业务流。同一个流中的所有分组具有相同的流标签，以便对有同样 QoS 要求的流进行快速、相同的处理。

目前，IPv6 的流标签定义还未完善，但从其定义的规范框架来看，IPv6 流标签提出的支持服务质量保证的最低要求是标记流，即给流打标签。流标签应该由流的发起者信源节点来赋予，同时要求在通信路径上的节点都能够识别该流的标签，并根据流标签来调度流的转发优先级算法。这样的定义可以使物联网节点上的特定应用有更大的调整自身数据流的自由度，节点可以只在必要的时候选择符合应用需要的服务质量等级，并为该数据流打上一致的标记。在重要数据转发完成后，即使通信没有结束节点也可以释放该流标记。这样的机制再结合动态服务质量申请和认证、计费的机制，就可以做到使网络按应用的需要来分配服务质量。同时，为了防止节点在释放流标签后又误用该流标签，造成计费上的问题，信源节点必须保证在 120s 内不再使用释放了的流标签。

在物联网应用中普遍存在节点数量多、通信流量突发性强的特点。与 IPv4 相比，由于 IPv6 的流标签有 20bit，足够标记大量节点的数据流。同时，与 IPv4 中通过五元组（源、目的 IP 地址，源、目的端口、协议号）不同，IPv6 可以在一个通信过程中（五元组没有变化），只在必要的时候才会赋予数据分组流标签，即在节点发送重要数据时，动态提高应用的服务质量等级，做到对服务质量的精细化控制。

当然，IPv6 的 QoS 特性并不完善，由于使用的流标签位于 IPv6 分组头，容易被伪造，产生服务盗用的安全问题。因此，在 IPv6 中，流标签的应用需要开发相应的认证、加密机制。同时为了避免流标签使用过程中发生冲突，还要增加源节点的流标签使用控制的机制，保证在流标签使用过程中不会被误用。

（4）IPv6 在物联网安全中的优势

物联网节点的安全性和可靠性也需要重新考虑。由于物联网节点限于成本约束，很多都只是基于简单的硬件，不可能处理复杂的应用层加密算法。同时，单节点的可靠性也不可能做得很高，其可靠性主要还是依靠多节点冗余来保证。因此，靠传统的应用层加密技术和网

络冗余技术很难满足物联网的需求。

IPv6 在网络安全方面有了更好的保障。由于物联网应用中节点部署的方式比较复杂，节点可能通过有线方式或无线方式连接到网络。因此节点的安全保障的情况也比较复杂。在使用 IPv4 的场景中一个黑客可能通过在网络中扫描主机 IPv4 地址的方式来发现节点，并寻找相应的漏洞。而在 IPv6 场景中，由于同一个子网支持的节点数量极大（达到百亿亿数量级），黑客通过扫描的方式找到主机的难度大大增加。在基础协议栈的设计方面，IPv6 将 IPSec 协议嵌入到基础的协议栈中，通信的两端可以启用 IPSec 加密通信的信息和通信的过程。网络中的黑客将不能采用中间人攻击的方法对通信过程进行破坏或劫持。即使黑客截取了节点的通信数据分组，也会因为无法解码而不能窃取通信节点的信息。

同时，由于 IP 地址的分段设计将用户信息与网络信息分离，使用户在网络中的实时定位很容易。这也保证了在网络中可以对黑客行为进行实时的监控，提升了网络的监控能力。

另一方面，物联网应用中由于成本限制，节点通常比较简单，节点的可靠性也不可能做得太高。因此，物联网的可靠性要靠节点之间的互相冗余来实现。又因为节点不可能实现较复杂的冗余算法，因此，一种较理想的冗余实现方式是采用网络侧的任播技术来实现节点之间的冗余。采用 IPv6 的任播技术后，多个节点采用相同的 IPv6 任播地址（任播地址在 IPv6 中有特殊定义）。在通信过程中发往任播地址的数据分组将被发往由该地址标识的“最近”的一个网络接口。其中，“最近”的含义指的是在路由器中该节点的路由矢量计算值最小。当一个“最近”节点发生故障时，网络侧的路由设备将会发现该节点的路由矢量不再是“最近”的，从而会将后续的通信流量转发到其他的节点。这样，物联网的节点之间就自动实现了冗余保护的功能。而节点上基本不需要增加算法，只需要应答路由设备的路由查询，并返回简单信息给路由设备即可。

IPv6 具有很多适合物联网大规模应用的特性。虽然 IPv6 还有众多的技术细节需要完善，但从整体来看，使用 IPv6 不仅能够满足物联网的地址需求，同时还能满足物联网对节点移动性、节点冗余、基于流的服务质量保障的需求，最终成为物联网应用的基础网络技术。

2．IPv6 在支持物联网方面存在的问题

IPv6 作为下一代 IP 网络协议，具有丰富的地址资源，能够支持动态路由机制，可以满足物联网对网络通信在地址、网络自组织以及扩展性方面的要求。但目前也存在一些技术问题需要解决。例如，无状态地址分配中的安全性问题，移动 IPv6 中的绑定缓冲安全更新问题，流标签的安全防护，全球任播技术的研究等。

另外，由于 IPv6 协议栈过于庞大复杂，不能直接应用到传感器设备中，需要对 IPv6 协议栈和路由机制作相应的精简，才能满足低功耗、低存储容量和低传送速率的要求。

由于 IPv6 协议栈过于庞大复杂，并不匹配物联网中互联对象（尤其是智能小物体）的特点。因此，虽然 IPv6 可为每一个传感器分配一个独立的 IP 地址，但传感器网需要和外网之间进行一次转换，起到 IP 地址压缩和简化翻译的功能。

目前，相关标准化组织已开始积极推动精简 IPv6 协议栈的工作。例如，IETF 已成立了 6LowPAN 和 RoLL 两个工作组进行相关技术标准的研究工作。相比较传统方式，IPv6 能支持更大的节点组网，但对传感器节点功耗、存储、处理器能力要求更高，因而成本要更高。另外，目前基于 IEEE 802.15.4 的网络射频芯片还有待进一步的开发来支持精简 IPv6 协议栈。

3．物联网对于 IPv6 的意义

物联网与 IPv6 各自的特点决定了它们之间必将产生紧密的联系，物联网构想的完全实现有赖于 IPv6 提供强大的支持。另一方面，物联网的普及必然会对 IPv6 的发展产生巨大的推动作用。

作为下一代互联网 IP 协议的 IPv6 取代 IPv4 是一个必然的趋势。但由于 NAT（网络地址翻译）和 DHCP 等技术的出现，暂时缓解了 IPv4 地址资源紧缺的现状，人们对于 IPv6 的需求变得相对不够强烈。需求动力不足加上人们对现有 IPv4 的依赖，导致 IPv6 取代 IPv4 的过程艰难而缓慢。

然而，前面我们已经分析过，物联网的发展对于 IPv6 产生了很大的需求，可以说，IPv6 在此看到了新的发展契机。为加快 IPv6 取代 IPv4 的进程，IPv6 的发展必须以物联网的发展为依托，从物联网的发展需求中获得自身发展的强大动力。由此我们也可以想象，除现在已有的一些应用外，IPv6 将来必定会首先被大量应用于新建的物联网系统中。随着物联网的不断发展壮大，IPv6 的应用范围也必将越来越广泛，最终超过 IPv4 的规模，形成对 IPv4 的相对优势，进而逐步淘汰并取代原有的 IPv4 协议。因此，物联网的发展对于 IPv6 的发展具有重要的驱动作用。

6.4　内容和应用迁移策略建议

随着互联网深入人类生活的各个角落，各种应用层出不穷。因此，对于互联网内容和应用向下一代互联网迁移不可能一蹴而就。从硬件基础设施层面来看，应逐步推进双栈 IDC 的建设，IDC 各类网络设备、服务器等网元均应支持双栈，具备同时提供 IPv4、IPv6 服务的能力。从应用软件层面来看，对于与用户直接交互的相关系统，首先升级支持双栈，后台系统升级，支持双栈业务的调度和管理；对于新增应用（如物联网）在建设初期就应考虑对 IPv6 的支持；对于原有应用的升级调整，也应把对 IPv6 支持的需求放在首位。通过以上措施，内容和应用将逐步实现向下一代互联网的迁移，为用户提供更好、更优质的服务。

6.4.1　大中型 ICP 的升级双栈改造

从上述分析可以看出，各类应用向 IPv6 迁移的难易程度不同。而对于大型 ICP 而言，往往是综合应用服务提供商，改造难度较大。首先，需将自身直接与用户终端交互的网元设备进行双栈化改造。其次，应用软件升级支持对 IPv6 用户的识别和管理，特别是对于提供的第三方链接等业务，还应提供 IPv6 缓存服务，避免纯 IPv6 用户无法进行访问。

6.4.2　小型 ICP 双栈的升级改造

一般情况下，小型 ICP 业务类型较为单一，可根据自身业务运营的实际情况，逐步向 IPv6 进行迁移。过渡技术部署初期，IPv6 业务量较小，可借助运营商提供的协议转换方案，进行 IPv6 内容和应用服务的迁移。随着 IPv6 用户的逐步增多和业务运营的不断发展，新增应用直接部署到支持双栈的 IDC 机房，并保证应用软件支持双栈，同时逐步迁移业务量增长较大原有应用支持双栈，实现业务向 IPv6 迁移的平稳过渡。

第 7 章
终端演进

7.1 当前各种终端支持 IPv6 能力现状

业务平台和 IT 支撑系统的大多数硬件和操作系统支持 IPv6，但软件不支持。

虽然目前互联网地址压力巨大，但向 IPv6 推进将会比较漫长。目前终端、应用和业务对于 IPv6 支持不足，用户软、硬终端（PC 除外）绝大多数不具备支持 IPv6 的能力。而终端作为距离用户端最近的设备，具有产业链长（涉及芯片、操作系统、终端设备制造等各个环节）、种类与数量众多等特性，在推动 IPv6 发展过程中起着至关重要的作用。

目前移动终端对于 IPv6 支持情况不太理想，国家尚无硬性规划，这直接影响 IPv6 发展的进程。GSM/TD 移动终端的 IPv6 技术要求的行业标准尚未出台，移动终端入网不对 IPv6 做强制要求，导致移动终端对 IPv6 支持薄弱。现网 2G 终端还有很多不支持 IPv6 的，TD 商用终端不支持 IPv6，中国移动仅有一款支持 IPv6 的 TD 终端原型与 TD 芯片。

终端企业迟迟不转向 IPv6 的其中一个重要原因是还没看到市场需求。运营商还没有提供 IPv6 商用的网络，终端企业即使开发了 IPv6 终端也无用武之地。这又归结到了产业链互相观望的死循环中。可喜的是，随着中国电信等运营商发布 IPv6 商用计划，部分领先的终端企业已经提出开发 IPv6 终端的计划，电信运营商也从部分终端企业定制了少量 IPv6 终端，这将有助于加快终端企业开发 IPv6 终端的进度。

中国电信于 2010 年年初正式启动下一代互联网部署试点工作，在长沙等 6 个城市开展 IPv6 试点。2011 年 5 月 31 日中国电信开通了纯 IPv6 网址的中国电信官网，并自主研发了面向 ICP 网站 IPv6 平滑迁移平台——“Smart6”，一使现有 ICP 网站具备 IPv6 服务能力，并为全球的 IPv6 用户提供服务。根据中国电信的规划，2009 年至 2010 年为 IPv6 的试商用阶段，2012 年至 2015 年规模商用 IPv6，2015 年以后将大规模使用 IPv6。

7.2 固网终端

对运营商来说，现网的复杂性，端、管、云三端的不同步性，决定了 IPv6 部署演进将是一个长期复杂的过程。IPv4、IPv6 地址在相当长时间内共存，地址复用机制将伴随整个迁移过程。

运营商建设 IPv6 网络，基本可分为试商用、规模商用、全面商用 3 个阶段。在该过程中，

接入终端也将按照这个路线进行技术选择，按照双栈引入、DS-Lite 逐步切换、6RD 局部补充的步骤来支撑 IPv6 的部署，同时支撑电信未来业务的运营，逐步演进到纯 IPv6 网络。

7.2.1　IPv6 对固网终端的技术要求

对于桥接型 CPE，对 IPv6 无额外要求，以下要求都是针对路由型 CPE 的。

1．IPv6 接入的模式

（1）IPv6 over PPP 方式

家庭网关支持 IPv6 PPP 拨号，通过路由方式为用户设备提供连接性。家庭网关工作在桥接模式下，每一个终端或主机发起各自的 PPPv6 拨号连接。

（2）IPv6 over Ethernet

家庭网关通过 IPoE 方式支持 IPv6 协议，为家庭网关获得地址及配置。

2．IPv6 接入的地址类型

（1）IPv6 的地址类型（IETF RFC 4291）

IPv6 的地址类型包括：单播地址、任播地址、多播地址。

（2）单播地址类型

IPv6 节点通信至少需要一个单播地址，单播地址类型及示例如下：

链路本地地址 link-local　　FE80::/10

唯一本地地址 unique-local　FC00::/7

全局单播地址（GUA）

（3）多播地址

① 基础 ND 协议使用所有节点多播地址 FF02::1。

② Solicited-node 多播地址为 FF02:0:0:0:0:1:FF + 接口 ID 后 24bit。

3．IPv6 地址分配方式

IPv6 的地址分配方式主要有手动配置、算法生成、用于分配地址前缀的 DHCP-PD、有状态的 DHCPv6 协议、无状态的地址自动配置。

4．DHCPv6-PD 过程

① BNG 作为指定路由器（DR），家庭网关作为申请路由器（RR）。

② 家庭网关发出带有 IA_PD 的 DHCPv6 消息。

③ BNG 为不同用户线路的家庭网关分配不同的前缀。

④ 不同用户线路决不能分配相同的地址前缀。

⑤ BNG 分配的地址前缀最大为/64，推荐为/56。

⑥ 家庭网关再从其中提取/64 的地址前缀分配给不同的 LAN 侧网段。

如图 7-1 所示。

5．DHCPv6 过程

① BNG 根据地址分配策略分配 128bit 的 IPv6 地址给 DHCPv6 客户端。

② 通过选项进行配置。

如图 7-2 所示。

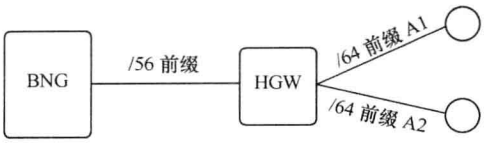


图 7-1　DHCPv6-PD 过程

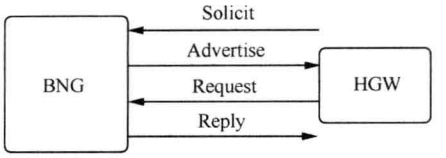


图 7-2　DHCPv6 过程

6. SLAAC 过程

① 由家庭网关自身的接口标识和路由器通告信息共同组成主机的 IPv6 地址。

② 路由通告（RA）信息的前缀信息，最多不要超过 64bit。

③ 接口标识根据一定规则生成，在子网内部具有唯一性。

④ 根据 48bit 的 MAC 地址生成 64bit 的接口 ID。

如图 7-3 所示。

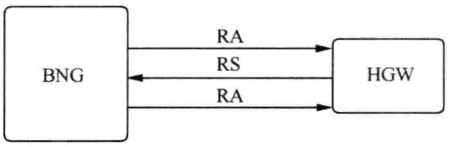


图 7-3　SLAAC 过程

7. PPP 的地址分配过程

① IPv6 的网络控制协议 IPv6CP 仅定义了 Interface-ID 的配置选项，生成该条线路的链路本地地址。

② 通过 PPP 分配 IPv6 的 GUA 过程建议如下。

a. 通过 PPP 获得的 64bit 的接口 ID 和通过 ND 协议获取的 64bit 前缀组成 IPv6 地址，通过 ND 协议扩展选项分配 DNS 等信息。

b. 通过 PPP 获得 64bit 的接口 ID，通过 ND 协议获取 64bit 前缀，组成 IPv6 地址，通过 DHCPv6 分配 DNS 等信息。

c. 通过 PPP 获得 64bit 的接口 ID，生成 link_local 地址，GUA 通过 DHCPv6 或 DHCPv6-PD 生成。

d. 通过扩展 IPv6CP 配置选项直接分配 128bit 地址、DNS 信息。

8. VLAN 和 QoS 要求

（1）支持根据以太网网帧类型划分不同 VLAN 的能力

☆ IPv6oE　0x86DD

☆ PPPoE　0x8863 或 0x8864

☆ IPv4oE　0x0800

☆ ARP　0x0806

（2）对上行流依据 IPv6 的特性进行分类或优先级标识

☆ IPv6 源地址

☆ IPv6 目的地址

☆ IPv6 Traffic Class

☆ IPv6 的 Flow Label

☆ 协议类型

☆ 端口

9. 安全要求

① 支持侦听上下行 DHCPv6 协议消息，建立 IA_NA 选项中的 IPv6 地址和 IA_PD 中的前缀与用户端口之间的映射关系。

② 支持侦听下行 RA 消息，并建立地址前缀与用户端口的映射关系。
③ 支持映射表的老化处理。
④ 支持侦听用户上行的重复地址检测分组 DAD，建立 IPv6 地址与用户端口的映射关系。
⑤ 限制 IPv6 的接入用户数目，限制使用的地址范围。
⑥ 分组过滤和转发。

10．邻居发现（ND）过程的支持

① 路由器请求（RS）。
② 路由通告（RA）。
③ 支邻居请求（NS）。
④ 支邻居通告（NA）。

11．端口定位

① LDRA 功能要求支持 option18（Interface-ID）的插入，用于标识用户线路。
② LDRA 功能要求支持 option37（Remote-ID）的插入，用于标识用户线路上的其他属性。
③ LDRA 功能要求支持 option17（Vendor-specific information）的插入。
④ LDRA 功能具备基于 VLAN 的进行开启和关闭的能力。

12．内置的 IPv6 主机支持

① 用于支持 IPv6 的 TR-069 管理。
② 用于提供一些主机业务。
③ 用于支持 IPv6 的 VoIP。

13．IPv6 多播

① 支持 MLDv1/v2 的多播数据流的转发，支持多播管理协议的 MIB。
② 支持 MLD Snooping 的透明功能和代理功能。
③ 支持其他多播数据流的处理功能，如加入、跳数限制、复制等。

14．管理功能的要求

① 支持基于 IPv6 的管理网络。
② 支持相关 IPv6 参数的配置管理。

7.2.2　IPv4/IPv6 过渡技术的演进策略

IPv6 是针对 IPv4 地址耗尽问题的根本解决方案。但是，因为 IPv6 与 IPv4 技术不兼容，导致从 IPv4 向 IPv6 难以实现平滑迁移。为了保证用户和业务平稳过渡，整个演进周期将是长期的过程，IPv6 会与 IPv4 长期共存。IPv6 的过渡，按照双栈引入、DS-Lite 逐步切换、6RD 局部补充的步骤依次进行。

1．双栈方案是 IPv6 试点初期的必然方案

双栈技术是指从用户侧到接入网、BRAS 同时支持 IPv4 和 IPv6 协议栈，如图 7-4 所示。双栈节点同时支持与 IPv4 和 IPv6 节点的通信。当和 IPv4 节点通信时需要采用 IPv4 协议栈，当和 IPv6 节点通信时需要采用 IPv6 协议栈。

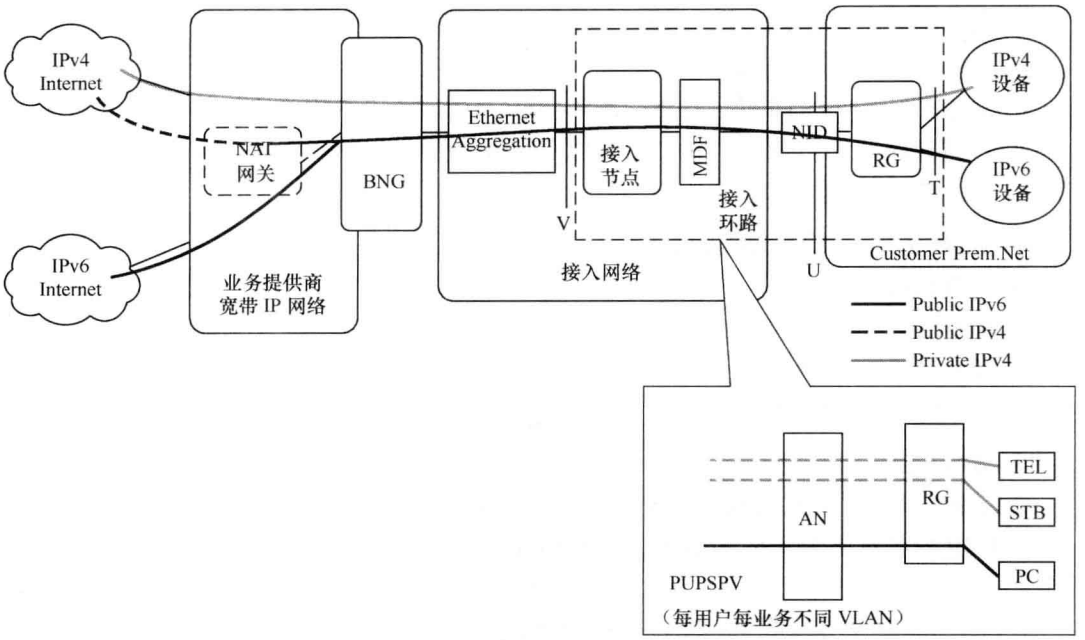


图 7-4　双栈技术

双栈解决方案是早期实验以及运营商发展 IPv6 初期的一个必然阶段。目的是验证 IPv6 技术本身承载业务的成熟性。处于双栈阶段，运营商的重要工作不仅仅是对网络设备的升级，还必须包括对 ICP 业务的牵引工作。回顾过去几年，运营商发展 IPv6 的顾虑在于，目前 IPv6 网络中的业务缺乏，用户迁移到纯 IPv6 网络直接面临网络应用下降，从而导致满意度下降的问题。相比国内其他几个运营商，中国电信拥有的不仅仅是最大规模的宽带接入用户，还包括了国内最大的 ICP 数量。因此，建议中国电信在考虑用户侧的同时，逐步迁移 ICP 到 IPv6，或者为 ICP 准备部分 IPv4、IPv6 双栈接入地址。早期的 IPv6 的 ICP 相对较少，往往能更容易地吸收用户流量。网络中的 DNS 也同步升级到 IPv6，为 IPv6 业务铺路。

但双栈方案并不节约 IPv4 地址，而只是减少新发展的业务对 IPv4 公网地址的消耗。随着客户端计算机的增加，运营商的 IPv4 地址仍然不够。同时，该方案并不解决 IPv4 与 IPv6 应用互访问题，所有流量都在封闭的网络中。如果这样发展下去，IPv4 与 IPv6 两张网络实际上将长期并存，相互不访问。虽然 IPv6 作为新增资源补充进来，但主要目的是把新业务放到 IPv6 网络中，而尽量延缓 IPv4 的使用时间。由于目前 IPv4 的主机和服务都占绝对优势，如果不考虑 IPv6 对 IPv4 网络的渗透，实际带来的结果就是网络流量大头还是在 IPv4 网络。

在 IPv6 部署初期，多重业务中的 Internet 业务首先以 IPv6 来解决。目前 VoIP、IPTV 都在单独的 VLAN 中，这部分不去迁移。Internet 业务迁移到 IPv6，并采用 NAT 方式实现与 IPv4 地址互通。运营商未来新封闭业务，如移动互联网，在 IPv6 中部署。终端 IPv4/IPv6 地址由 BNG 采用双栈方式分别分配。

终端需实现双栈技术。IPv4/IPv6 数据流跑在不同 VLAN 中，终端可按照端口不同，区别 IPv4/IPv6 数据流；终端也可以使用桥接技术简化 IPv6 实现难度。接入网在 IPv4/IPv6 下提供一致转发，相比之下，IPv6 VLAN 下的管理能力更为重要（面向公众的开放业务）。

2. DS-Lite 方案具有较强的生命力

双栈方案并不能解决IPv4即将耗尽的问题，解决地址耗尽只能靠复用IP地址。传统的NAT44方式只能是业务长期固守于 IPv4 上，时间一长还是会碰到地址耗尽和迁移到 IPv6 的问题。

DS-Lite 是 4over6 隧道 + NAT44 的组合技术，如图 7-5 所示。双栈主机采用 IPv6/IPv4 地址和 IPv6/IPv4 节点进行通信。其中，终端对 IPv6 流简单路由转发，对于 IPv4 流以与 CGN 之间的 4over6 隧道方式承载。CGN 设备将 IPv4 报文剥离，通过 NAT44 技术转换成公网地址访问 IPv4 节点。

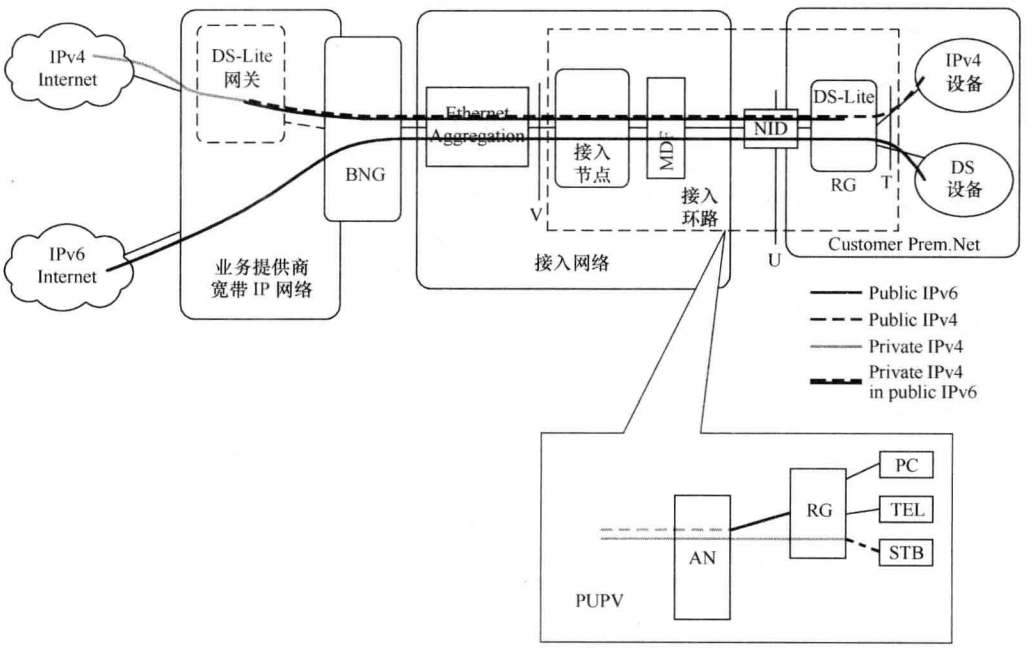


图 7-5　DS-Lite 技术

DS-Lite 方案比较巧妙地结合了 NAT44 和 IPv6 双栈这两种思路。考虑到 IPv4 与 IPv6 会在相当长的时间共存，主机以 IPv4 单栈和 IPv4/IPv6 双栈为主的实际情况，DS-Lite 采用 NAT44 方式进行 IPv4 传统业务的访问。而同时网络逐步进行 IPv6 化改造，这样网络升级与传统业务的支持就完全独立开，运营商可以直接升级其业务网络。对于一个产业链各个节点都需要升级的演进过程，这一点非常重要。

在 DS-Lite 的建设过程中，运营商升级网络（包括城域网、接入网）到支持 IPv6（或者 IPv4/IPv6 双栈），发放支持 DS-Lite 的 RG。虽然先期投入较大，但整个网络平滑演进。另外，DS-Lite 的 RG 发放，可以与运营商现有家庭网关的发放工作结合起来。

ICP 的建设需要保持双栈的思路。这不仅仅是业务迁移的需要，同时也是 DS-Lite 实现业务所必须的。DS-Lite 并不提供 IPv4 单栈主机对 IPv6 服务器访问的支持，所以对开展在 IPv6 网络中的新业务，还必须提供 IPv4 的访问机制。也就是说，业务系统及其所在网络必须支持双栈，分别对来自 IPv4 和 IPv6 的主机服务。但这点对目前的运营商网络并不是太大问题，只是会增加业务系统的复杂度。

在该阶段，运营商网络将一步切换到 IPv6，将 IPv4 的流量逐步吸收。支持 IPv4 单栈的计算机访问 IPv4 网络，DS-Lite 终端与 DS-Lite 网关配合实现单播业务 IPv4 over IPv6。

接入网感知 IPv6 流量，提供 IPv6 单播。IPv6 地址由 BNG 分配，IPv4 地址由 RG 分配私网地址。多播业务可切换到 IPv6，或者保留在原来的 IPv4 VLAN 中。终端支持 IPv6 DS-Lite、IPv6 路由功能。

3. 局部地区以 6RD 方案作为补充

无论是双栈还是 DS-Lite 方案，都需要从客户终端到 BRAS 的整个接入网都进行改造，运营商投资很大。对于某些运营商来说，他们不愿意立刻就把所有的现网设备进行升级，或者局部地区暂时没有改造条件，这时候可以用支持 6RD 的终端（RG 或者在 PC 上安装软件），接入 IPv6 网络。

6RD 是 IPv6 快速部署（IPv6 Rapid Deployment）的简称，是在 6to4 基础上发展起来的一种 IPv6 网络过渡技术方案，如图 7-6 所示。6RD 通过在现有 IPv4 网络中增加 6RD-BR，给愿意使用 IPv6 的用户提供 IPv6 接入，在 IPv6 用户的家庭网关和 6RD 网关之间建立 6in4 隧道，从而实现在 IPv4 网络提供 IPv6 服务的能力。

但 6RD 的解决方案还依赖于 IPv4 地址，对于 IPv6 地址的计算，需要 IPv4 地址。虽然在一个家庭内部，多个主机可以共享这个 IPv6 的前缀，从而达到降低 IPv4 地址使用的能力，但其目的重点也不是解决 IPv4 地址问题，而只是解决不升级网络，分配 IPv6 地址的问题。因此，6RD 是解决不准备迁移到 IPv6 的运营商，为 IPv4 用户提供 IPv6 服务的问题。只能作为部分地区的补充，不是向 IPv6 迁移的方案。

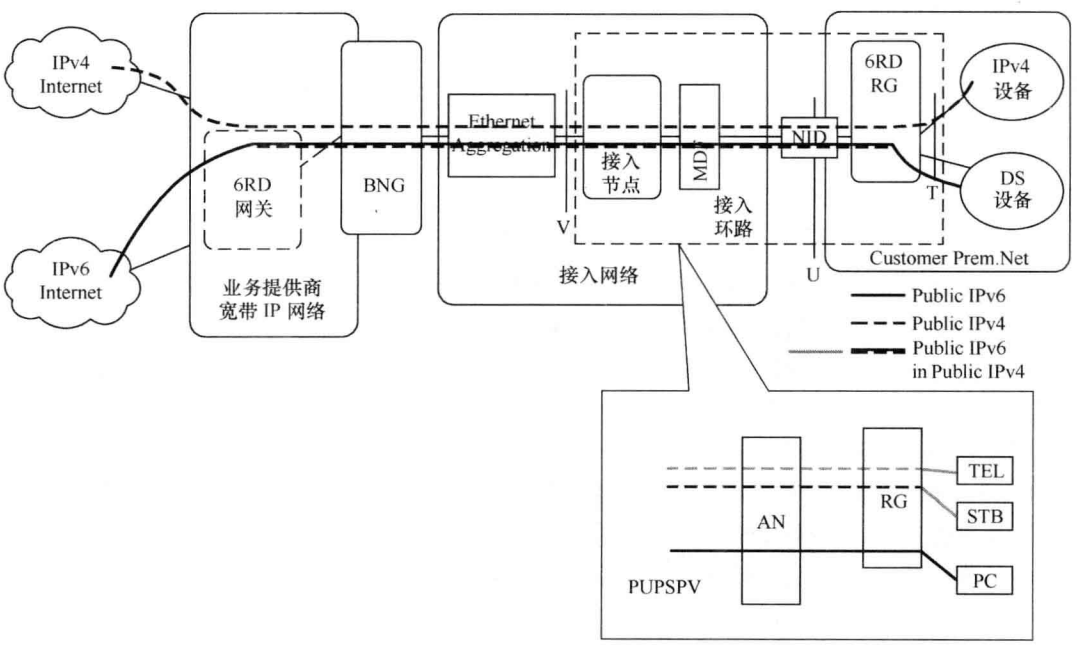


图 7-6　6RD 技术

7.3　PC 操作系统和客户端

目前，各种常见的操作系统对 IPv6 的基本功能有不同程度的支持。微软公司从 Windows XP 开始已经可以支持 IPv6 协议栈的基本功能，新一代的操作系统 Windows 7 和 Windows

Server 2008 对 IPv6 的支持相对于之前的 Windows XP 系列有了较大的提升。从 Windows Vista 开始，操作系统就默认安装并开启 IPv6 协议栈，而且支持通过图形化用户界面对 IPv6 协议栈的网络参数进行配置。同时，Windows 7 系列还支持 DHCPv6 动态自动配置，使得 IPv6 的部署更加灵活。

对于名称解析功能，Windows XP 系列在双栈情况下默认通过 IPv4 的 DNS 服务器进行域名解析，而且不支持 LLMNR 和 ipv6.literal.net 域名。

IPv6 的连接能力是直接影响到 IPv6 普及的重要因素。Windows XP 目前仍不支持 IPv6 的 PPPoE 拨号，需要安装第三方软件才能实现。而 Windows 7 提供了官方的 PPPoE 拨号程序，使得 Windows 7 系列将会在 IPv6 环境下更加稳定。此外，Windows XP 在其他 IPv6 连接能力的支持上都不是很理想。

基于开放源码的 Linux 对 IPv6 提供了比较强的支持。Linux 内核中大多数 IPv6 功能都是来自于 USAGI 项目（Universal playground for IPv6），在内核中实现的都是 USAGI 的核心功能。Linux 在它的基础系统中集成了一个支持 IPv6 的 PPP 服务来提供基本的 IPv6 功能。

7.4　手机终端

手机终端对 IPv6 的支持依赖于采用的主芯片平台，对于 CDMA 终端产品来说，主要的芯片平台有两个，一个是高通平台，另外一个是威盛平台，其中高通平台占绝对主流的份额。

高通主流平台对 IPv6 的支持进度如图 7-7 所示，后续新的平台基本上都支持了 IPv6，可以配合电信运营商网络进行相应的测试和商用。

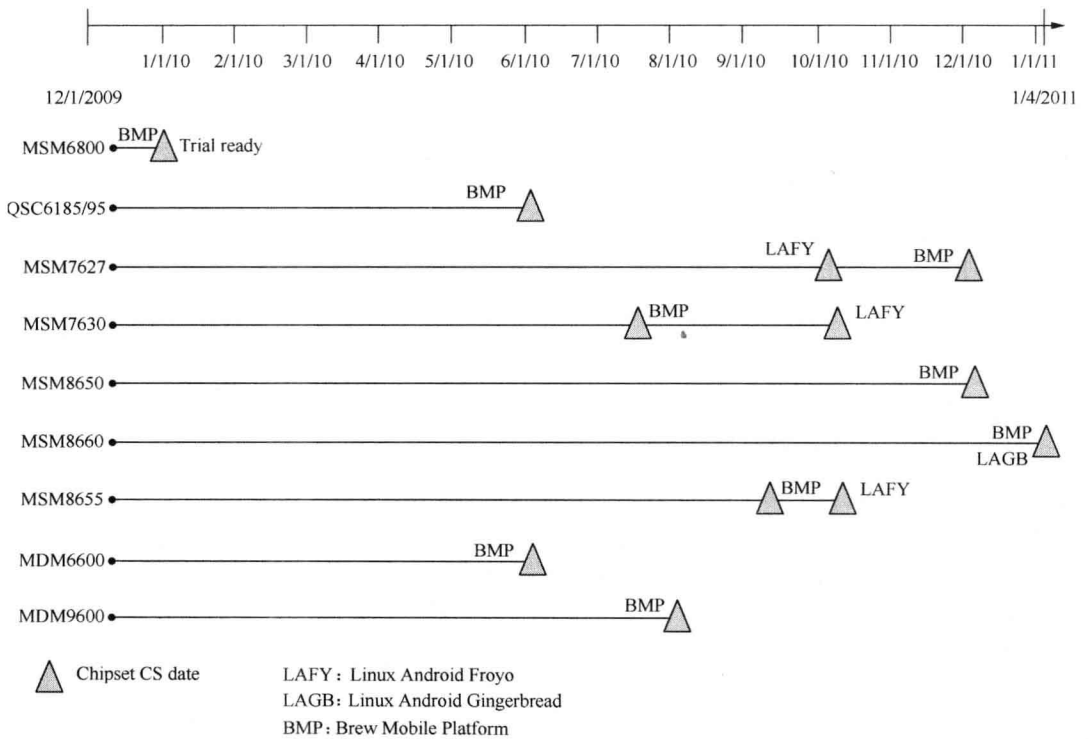


图 7-7　高通平台的 IPv6 支持情况

威盛公司的 EVDO 芯片 CBP7.X 支持 IPv6/IPv4 双栈，CBP7.X 已经商用。

IPv6 在智能和非智能手机终端的支持要求是不一样的。

对于智能终端，初期可以采用 Relay 模式来支持 IPv6/IPv4 双栈。智能手机终端需要提供支持 IPv6 和 IPv4 双栈拨号的驱动程序。智能手机终端所使用的操作系统需要支持 IPv6/IPv4 双栈。逐步推进智能手机终端支持 QoS 功能。后期智能手机终端逐步采用 AT 增强模式来支持 IPv6/IPv4 双栈。

非智能手机终端做 Modem 使用时，需要提供支持 IPv6 和 IPv4 双栈拨号的驱动程序。

7.5 3G 上网卡

目前的3G上网卡业务中，用户通过PC终端上的WiFi网卡和3G上网卡以及无线宽带UIM卡，使用无线宽带客户端软件或认证门户，选择 WLAN、3G、1x 等不同网络接入方式上网。

与固网宽带的家庭网关类似，3G 上网卡分为路由模式和桥接（Relay）模式两种，两种模式下对 IPv6 支持的要求是不同的。

路由模式下的上网卡要求芯片支持 IPv6 协议栈，这与手机终端中芯片对 IPv6 的支持要求类似。目前芯片支持能力较弱，因此市场上还没有支持 IPv6 的路由型 3G 上网卡。

桥接模式下的上网卡是一个纯粹的数据通道，IPv6 特性的支持完全依赖于 PC 的操作系统实现。以 Windows 操作系统为例，由于 Windows XP 系统对于 IPv6 的支持尚不完善，需要通过定制的无线宽带客户端软件才能实现 IPv6 拨号功能。而 Window 7 因为已经具备了官方的 PPPoE 拨号程序则能较好地支持 IPv6 拨号功能。

7.6 IPTV 机顶盒

中兴 IPTV 机顶盒由 IPv4 向 IPv6 演进大致分为 3 个阶段。

1. 第一阶段：机顶盒和承载网的 IPv6/IPv4 双栈能力验证阶段

该阶段要求承载网支持 IPv6 多播。机顶盒支持 IPv6/IPv4 双栈，并支持 IPv6 多播业务管理平台，支持 IPv6 地址的管理，能力平台支持 IPv6 多播。相关的验证工作在试验网上进行。在该阶段，机顶盒的主要工作在于支持 IPv6 多播，如图 7-8 所示。

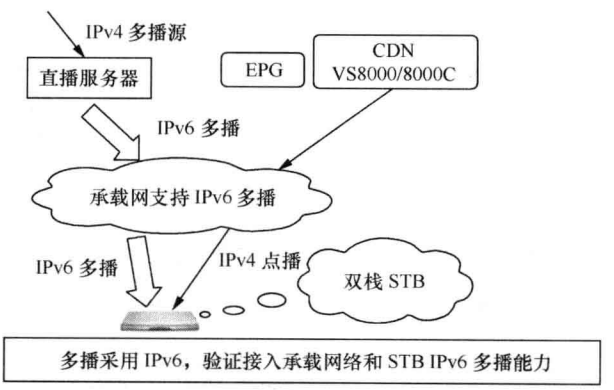


图 7-8 机顶盒和承载网的 IPv6/IPv4 双栈能力验证阶段

2. 第二阶段：IPv6/IPv4 兼容阶段

该阶段在第一阶段的基础上，通过软件升级，IPTV 中的各个网元全面支持 IPv6/IPv4 双栈。机顶盒除了支持 IPv6 多播外，其与 IPTV 系统的不同网元之间的通信也要求支持 IPv6，具体来说主要包括：机顶盒和认证管理平台，机顶盒和 CDN，机顶盒和 EPG，机顶盒和第三方增值业务平台，机顶盒和网管等。另外，

还包括：机顶盒和 NTP 服务器，机顶盒本地的配置管理工具等。具体情况如图 7-9 所示。

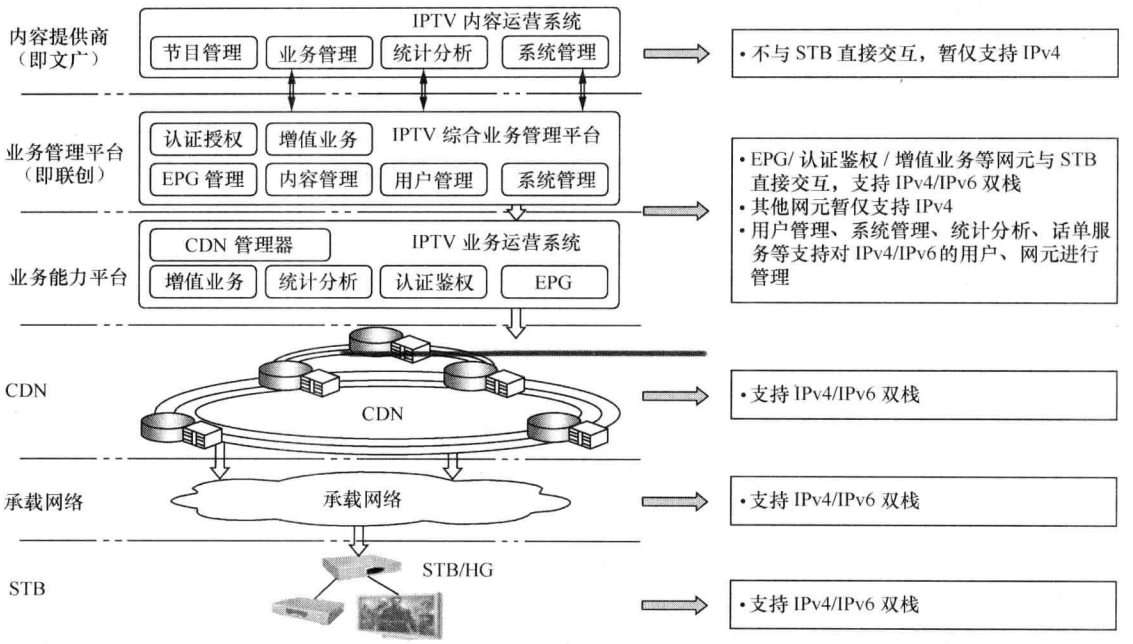


图 7-9　IPv6/IPv4 兼容阶段

3. 第三阶段：全网支持 IPv6 阶段

当机顶盒配置成纯 IPv6 工作模式时，机顶盒和 IPTV 系统中的各个网元使用纯 IPv6 的方式进行通信。当机顶盒工作在这种模式时，要求整个 IPTV 系统的各个网元已经全面支持 IPv6。

7.7　物联网终端

物联网终端是物联网中连接传感网络层和传输网络层，实现采集数据及向网络层发送数据的设备。它担负着数据采集、初步处理、加密、传输等多种功能。

物联网终端属于传感网络层和传输网络层的中间设备，也是物联网的关键设备。通过它的转换和采集，才能将各种外部感知数据汇集和处理，并将数据通过各种网络接口方式传输到互联网中。如果没有它的存在，传感数据将无法送到指定位置，“物”的联网将不复存在。

7.7.1　物联网终端的分类

1. 应用领域分类

主要包括工业设备检测终端、设施农业检测终端、物流 RFID（射频识别）终端、电力系统检测终端、安防视频监测终端、智能家庭终端等，下面就几个常用行业介绍一下终端的主要特点。

（1）工业设备检测终端

该类终端主要安装在工厂的大型设备上或工矿企业的大型运动机械上，用来采集位移传感器、位置传感器（GPS）、震动传感器、液位传感器、压力传感器、温度传感器等数据，通

过终端的有线网络或无线网络接口发送到中心处理平台，进行数据的汇总和处理，实现对工厂设备运行状态的及时跟踪和大型机械的状态确认，达到安全生产的目的。抗电磁干扰和防暴性是此类终端考虑的重点。

（2）设施农业检测终端

该终端一般被安放在设施农业的温室/大棚中，主要采集空气温湿度传感器、土壤温度传感器、土壤水分传感器、光照传感器、气体含量传感器的数据，将数据打包、压缩、加密后通过终端的有线网络或无线网络接口发送到中心处理平台进行数据的汇总和处理。这种系统可以及时发现农业生产中不利于农作物生长的环境因素，并在第一时间内通知使用者纠正这些因素，提高作物产量，减少病虫害发生的概率。终端的防腐、防潮设计将是此类终端的重点。

（3）智能家庭终端

该类设备一般应用于家庭内部的环境监测、家庭娱乐、家庭保健、家庭监控、家庭安防、家庭能源管理、家庭监护、紧急救助等，通过这些设备可以方便地实现家庭内部的智能管理和控制，达到出门在外、家就在身边的目的。

2．使用扩展性分类

物联网终端可以分为单一功能终端和通用智能终端。

单一功能终端一般外部接口较少，设计简单，仅满足单一应用或单一应用的部分扩展。除了这种应用外，在不经过硬件修改的情况下无法应用在其他场合中。目前市场上此类终端较多，如汽车监控用的图像传输服务终端、电力监测用的终端、物流用的 RFID 终端。这些终端的功能单一，仅适用在特定场合，不能随应用变化进行功能改造和扩充等。因为功能单一，所以该类终端的成本较低，也比较易于标准化。

通用智能终端因考虑到行业应用的通用性，其外部接口较多，设计复杂，能满足两种或更多场合的应用。它可以通过内部软件的设置、修改应用参数，或通过硬件模块的拆卸来满足不同的应用需求。该类模块一般涵盖了大部分应用对接口的需求，并具有网络连接的有线、无线等多种接口方式，还扩展了如 Bluetooth、WiFi、ZigBee 等接口，甚至预留一定的输出接口用于物联网应用中对“物”的控制等。该类终端一般能够接入数量较多的传感器，能够对传感数据进行汇聚、融合、转换、转发，以及对传感器进行较为复杂的配置和管理。物联网网关就是此类的通用智能终端。

7.7.2 物联网网关

物联网网关（M2M Gateway）也可称为网间连接器、协议转换器。物联网传感末梢节点类型丰富，使用环境复杂，在一个业务应用中其传感终端数量从几个到上万个。如果要求每个传感终端都做公网接入，会大幅提高传感终端的复杂度以及设备成本，在这种情况下，物联网网关应运而生。

网关是一种充当协议转换的计算机系统或嵌入式设备，可用于连接两种使用不同通信协议、数据格式或语言，甚至不同体系结构的异构系统，实现传输层以上的异构网络互联，是最复杂的网络互联设备。大多数网关运行在 OSI 第 7 层协议——应用层，实现应用层的异构网络互联，至少需要进行网络层、数据链路层及物理层的协议转换，十分复杂。网关对接收

的数据分组进行重新打包，以适配目的系统的特点，还可提供过滤和安全功能。网关实现异构网络互联，可以选择在传输层或者应用层上完成，并相应地对不同层次的协议进行转换。

（1）网关实现网络互联的层次

在传输层实现互联。当两个网络采用不同的传输层协议时，需要对传输层及以下各层进行转换，称为传输层网关。

在应用层上实现互联。两个异构网络使用不同的应用层协议时，网关实现应用层及以下各层的协议转换，即应用层网关。传感器网络中协议繁多，大多还使用专有的应用层协议，与互联网的应用层协议不同。因此，传感器网络网关通常要在应用层实现转换，属于应用层网关。

（2）网关的基本功能

① 多网络接口。网关设备与多种异构的网络相连，比如以太网、ZigBee、RS-485 总线等，网关设备必须提供多种硬件接口，与异构网关的设备相连。在实际应用中，网关设备通常提供 RS-232 接口、无线通信模块、以太网接口等多种类型的硬件接口。

② 协议转换功能。协议转换是网关最主要的功能。网关根据所连接的网络的类型，提供从物理层到运输层，甚至到应用层的协议转换功能。根据应用场合的不同特点，协议转换能力可以不同，比如，有的只需要物理层到运输层的协议转换，有的则需要完成物理层到应用层的协议转换。协议转换涉及地址转换、数据分组格式转换、应用层属性等映射和转换、协议交互机制的转换等多个层面，十分复杂。

③ 速率匹配和流量控制。当两个不同工作速率的网络进行互联时，网关需要将源网络速率转换到目的速率，并提供某种流量控制机制来限制输入到目的网络上的数据量。网关常用的流量控制方式有：源站仅在得到允许时才进行传输；强制源站减少向网络负荷；网关在网络拥塞时主动丢弃数据分组，并通知发送端。

④ 可靠数据传送功能。为了在互联的异构网络中可靠地传输数据，网关常采用以下一些措施：防止分组在若干个网关中无限制的循环；向源站或者其他网关发送错误报告；提供网间信息的重传功能。

⑤ 数据分组分片和组装能力。在异构网络中，由于采用的物理传输链路可能差异很大，支持的数据分组大小各不相同，因此网关需要对数据分组进行分片和组装。

⑥ 具有路由选择功能。网关设备连接了多个异构的网络，当接收到数据分组时，它需要选择通过哪个物理接口转发数据，必须具备路由功能。

另外，对于应用在不同领域的物联网网关，需要实现各自领域内的特别需求。

1．安防领域的物联网网关

安防领域的物联网网关除实现上述的基本功能外，还需要具备如下功能。

① 安全管理功能。安全管理对于安防应用至关重要，网关设备必须考虑对安全管理的支持。安全管理涉及加密、解密算法、密钥管理算法、认证和鉴权、持久化存储加密机制、设备准入控制机制等安全管理功能。由于这些算法对计算能力和存储能力要求很高，难以在传感器节点上实现，需要在网关中嵌入安全管理功能，实现传感器网络内部以及广域网络内的安全通信，并对设备进行认证和鉴权，防止非授权操作。

② 网络接口管理和诊断。网关通过多个物理接口与异构的多种网络相连。为了长时间可靠工作，当网络接口出现故障时，网关能够及时进行报警、维修，或者自动切换到备份系统，

确保可靠性。网关需要提供接口管理功能，自动对网络接口的运行状态进行检测和诊断。

③ 可靠备份。在互联网中，网关的作用主要是协议转换的功能，无需对连接的网络进行管理，无需存储网络的相关数据。而在传感器网络中，网关需要保存大量底层网络的信息，如节点 ID、节点状态、资源分配、密钥等。网关一旦出现故障，数据将被丢失，进而导致整个网络重新建立，降低了网络的性能和服务质量。因此，为了保障网络的可靠性，通常需要对网关进行热备份或者冷备份。

④ 远程管理和配置。在实际应用中，传感器网络布设在偏远不易触及的地方，到现场进行网络维护费时费力，代价很高。因此，需要在网关中运行网络管理程序，支持用户远程管理和配置网关及传感器网络。

⑤ 区分服务质量管理功能。网关可以连接面向多种应用的网络，数据类型以及对服务质量的需求不相同，如视频要求高带宽、音频对时延很敏感、报警数据要求及时可靠传输等。网关需要具备服务质量管理功能，根据数据类型进行相应的处理，确保满足应用中对服务质量的需求。

2．工业自动化领域的物联网网关

工业自动化领域内使用的物联网网关，一般要求具备高可靠性、实时性和高吞吐量，另外还需要具备如下功能。

① 安全管理功能。同样，由于传感器网络不依赖固定基础设施来对网络的安全进行管理，因此还需要在网关中嵌入安全管理功能，以防止内部妥协节点报告虚假信息，防止外部恶意节点通过制造冲突、复制延迟、改变信号强度等手段影响节点距离测量和跳数计算，进而影响网络性能。

② 时间同步功能。通常，网关还扮演着时钟源这一重要角色，为网络提供时间同步服务，以确保事件发生的先后次序。

③ 确定性调度功能。其主要功能是在相互竞争的用户之间分配通信资源，从而避免发生冲突，提高吞吐量和带宽利用率，解决现场设备之间数据通信的确定性、可靠性和实时性问题。通常，在网关中也需要完成该功能。

7.7.3 短距离无线通信技术

物联网网关以及各种物联网终端在传感器网络层较多采用短距离无线通信，例如采用 ZigBee、Bluetooth、WiFi 等技术，下面介绍几种短距离无线通信技术。

1．ZigBee

（1）技术介绍

ZigBee 技术是一种具有统一技术标准的短距离无线通信技术。它是为低速率控制网络设计的标准无线网络协议，依据 IEEE 802.15.4 标准，在数千个微小的传感器之间相互协调，实现通信。只需要很少的能量，就能以接力的方式通过无线电波将数据从一个节点传到另一个节点，从而实现在 2.4GHz 免费频带范围内的高效、低速率的通信功能。

ZigBee 设备具有能量检测和链路质量指示的功能，它采用了碰撞避免机制，以避免发送数据时产生数据冲突。在网络安全方面，ZigBee 设备采用了密钥长度为 128bit 的加密算法，对所传输的数据信息进行加密处理，从而保证了数据传输时的高可靠性和安全性。

目前 ZigBee 技术在国外已经在家庭网络、控制网络和手机移动终端等领域有了一定的应用，由 ZigBee 技术构建的无线传感器网络具有功耗低、成本低、结构简单、体积小、性价比高、扩展简便、安全可靠等显著特点。它是由一组 ZigBee 节点以 Ad hoc 方式构成的无线网络，主要采取协作方式，有效地感知、采集和处理网络覆盖范围内感知对象的信息。传感器网络中的部分或全部节点都是可以移动的，传感器网络的拓扑结构也会随着节点的移动而不断发生变化。每个传感器节点都具有动态搜索、定位跟踪和恢复连接的能力。因此，这种新兴的无线传感器网络技术必将有广泛的应用前景。图 7-10 为 ZigBee 无线传感器网络体系结构图。

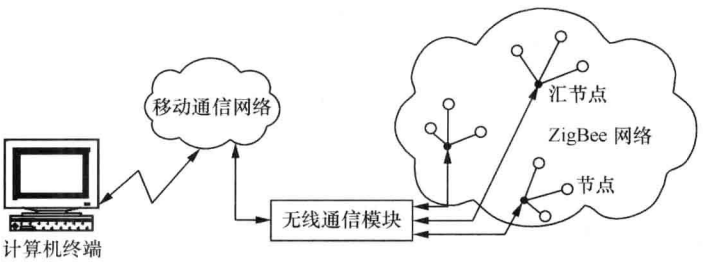


图 7-10　ZigBee 无线传感器网络体系结构

（2）基本参数

① 频带

使用的频段分别为 2.4GHz（全球通用）、868MHz（欧洲），以及 915MHz（美国），为免执照频段，带宽大约 3MHz。

② 距离及速率

数据传输的速率低（10～250kbit/s），有效覆盖范围一般为 10～75m，依据实际发射功率的大小以及具体应用来决定，基本上可以满足一般的办公室工作环境以及家庭使用。

③ 功耗及成本

在低耗电待机模式下，两节普通 5 号电池可以用 6 个月到 2 年。ZigBee 协议简单，ZigBee 协议免收专利费，成本较低。

（3）ZigBee 网络的组成

ZigBee 标准采用一整套技术来实现可扩展的、自组织的和自恢复的无线网络，并能够管理各种数据传输模式。为了降低系统成本，ZigBee 网络依据 IEEE 802.15.4 标准，定义了两种类型的物理设备，即全功能设备和简化功能设备。在 ZigBee 网络中每一个节点都具备一个无线电收发器、一个很小的微控制器和一个能源。这些装置将互相协调工作，以确保数据在网络内进行有效的传输。依据 IEEE 802.15.4 标准，ZigBee 网络将这两种物理设备在逻辑上又定义成为 3 类设备，即 ZigBee 协调器、ZigBee 路由器和 ZigBee 终端设备。

（4）ZigBee 网络的拓扑结构

ZigBee 网络层主要支持 3 种类型的拓扑结构，即星形（Start）网络结构、网状（Mesh）网结构和簇—树状（Cluster-Tree）网络结构。

① 星形网络结构

星形网络是由一个 ZigBee 协调点和一个或多个 ZigBee 终端节点构成的。ZigBee 协调点必须是 FFD，它位于网络的中心位置，负责发起建立和维护整个网络，其他的节点一般为 RFD

（也可以为 FFD），它们分布在 ZigBee 协调点的覆盖范围内，直接与 ZigBee 协调点进行通信。

② 网状网结构

网状网一般是由若干个 FFD 连接在一起组成的骨干网。它们之间是完全的对等通信，每一个节点都可以与在它的无线通信范围内的其他节点进行通信，但这些节点中也有一个会被推荐为 ZigBee 的协调点，如可以把第一个在信道中通信的节点作为 ZigBee 协调点。

③ 簇—树状网结构

簇—树状网结构中，节点可以采用 Cluster-Tree 路由来传输数据和控制信息。枝干末端的叶子节点一般为 RFD。每一个在它的覆盖范围中充当协调点的 FFD 向与它相连的节点提供同步服务，而这些协调点又受 ZigBee 协调点的控制。ZigBee 协调点比网络中的其他协调点具有更强的处理能力和存储空间。簇—树状网的一个显著优点就是它的网络覆盖范围非常大，但随着覆盖范围的不断增大，信息传输的时延也会逐渐变大，从而使同步变得越来越复杂。

2．Bluetooth

（1）技术介绍

蓝牙（Bluetooth）技术是由爱立信、诺基亚、Intel、IBM 和东芝 5 家公司于 1998 年 5 月共同提出开发的，并联合成立了蓝牙特殊利益小组（SIG），负责开发无线协议规范并设定交互操作的需求。蓝牙技术的本质是设备间的无线联接，主要用于通信与信息设备。蓝牙技术的支持者也很多，最初只有 5 家企业发起的 SIG 发展到现在已拥有了 3000 多个企业成员。

蓝牙技术是一种支持点对点或点对多点的话音、数据业务的短距离无线通信技术。它支持设备短距离通信（一般是 10m 之内）的无线电技术。蓝牙的标准是 IEEE 802.15，工作在 2.4GHz 频带，带宽为 1Mbit/s，能在包括移动电话、PDA、无线耳机、笔记本电脑、相关外设等众多设备之间进行无线信息交换。

截止到 2009 年 4 月，蓝牙共有 5 个版本 v1.1/1.2/2.0/2.1/3.0，其规范的技术和功能也逐渐完善。随着 Bluetooth 1.2 的推出，在传输速度、抗干扰、安全性等方面都有很大的提高。依据其发射输出的电频有 3 种距离等级，Class1 为 100m 左右、Class2 约为 10m、Class3 为 2～3m。一般情况下，其正常的工作范围是 10m 半径之内。2.0 版本的蓝牙标准提供的带宽将是目前蓝牙带宽的 3 倍。不过新版本蓝牙标准并没有增大无线传输的距离。蓝牙 2.0 在传输大量文件数据时耗电量只有旧版本蓝牙标准的一半。另外，新版本的蓝牙标准还具有向后兼容能力和更小的误码率。2009 年 4 月 21 日，蓝牙技术联盟（Bluetooth SIG）正式颁布了新一代标准规范——Bluetooth Core Specification Version 3.0 High Speed（蓝牙核心规范 3.0 版 高速）。

（2）基本参数

① 频带以及带宽

蓝牙的标准是 IEEE 802.15，工作在 2.4 GHz ISM 频带，带宽为 1MHz，跳频为 79 个频段。

② 传输距离以及速率

传输距离一般在 10m 左右，最远可达 100m，蓝牙的空中传输速率在 1～3Mbit/s，其中在 2009 年 4 月发布的 V3.0，通过集成 802.11 PAL（协议适应层），蓝牙 3.0 的数据传输速率提高到了约 24Mbit/s（即可在需要的时候调用 802.11 WiFi 用于实现高速数据传输），是蓝牙 2.0 的 8 倍。

③ 组网

蓝牙系统采用一种灵活的无基站的组网方式，使得一个蓝牙设备可同时与 7 个其他的蓝牙设备相连接。基于蓝牙技术的无线接入简称为 BLUEPAC（Bluetooth Public Access）。蓝牙系统的网络拓扑结构的有两种形式：微微网（Piconet）和分布式网络（Scatternet）。

微微网是通过蓝牙技术以特定方式连接起来的一种微型网络。一个微微网可以只是两台相连的设备，比如一台便携式电脑和一部移动电话，也可以是 8 台连在一起的设备。在一个微微网中，所有设备的级别是相同的，具有相同的权限。蓝牙采用自组式组网方式（Adhoc）。微微网由主设备（Master）单元（发起链接的设备）和从设备（Slave）单元构成，包括 1 个主设备单元和最多 7 个从设备单元。主设备单元负责提供时钟同步信号和跳频序列，从设备单元一般是受控同步的设备单元，接受主设备单元的控制，组网如图 7-11 所示。

在用蓝牙技术组建无线局域网的时候，组网的无线终端设备不超过 7 台，即组建一个微微网。组建方式有两种，一种是 PC 对 PC 组网，另一种是 PC 对蓝牙接入点组网。

在 PC 对 PC 组网模式中，一台 PC 通过有线网络接入 Internet 之中，利用蓝牙适配器充当 Internet 共享代理服务器。另外一台 PC 通过蓝牙适配器与代理服务器组建蓝牙无线网络，充当一个客户端，从而实现无线连接，共享网络的目的。这种方案是在家庭蓝牙技术组网中最具有代表性和最普遍采用的方案，具有很大的便捷性，组网如图 7-12 所示。

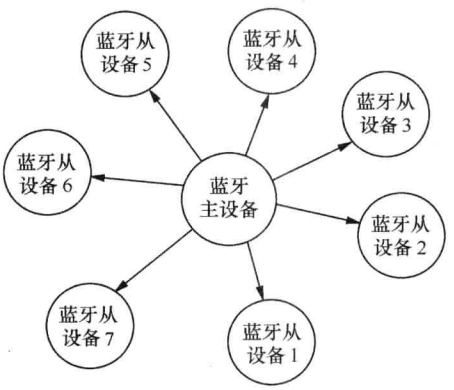


图 7-11　微微网组网图

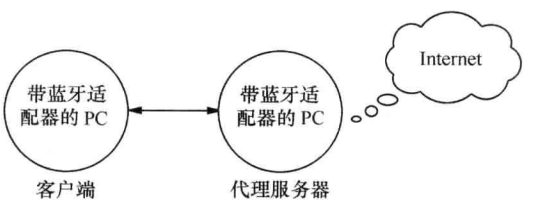


图 7-12　PC 对 PC 组网图

在 PC 对蓝牙接入点的组网模式中，蓝牙接入点（即蓝牙网关），通过与 Modem 等宽带接入设备相连接入 Internet。以蓝牙网关发射无线信号，与各个带有蓝牙适配器的终端设备相连接，从而组建一个无线网络，实现所有终端设备共享网络的目的。终端设备可以是 PC、笔记本电脑、PDA 等，但它们都必须带有蓝牙无线功能，且不能超过 7 台终端。这种方案适用于公司、企业组建无线办公系统，具有很大的便捷性和实用性，组网如图 7-13 所示。

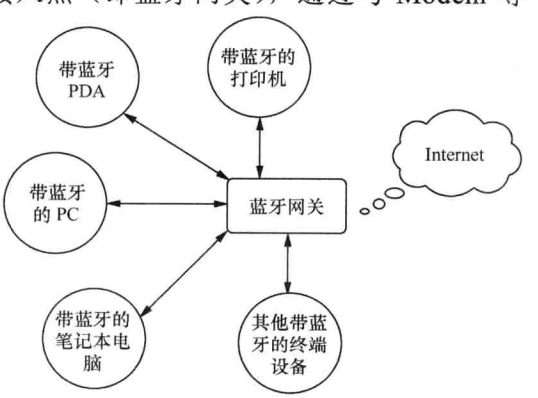


图 7-13　PC 对蓝牙接入点组网图

分布式网络是由多个独立的非同步的微

微网组成的，以特定的方式连接在一起。一个微微网中的主设备单元同时也可以作为另一个微微网中的从设备单元，这种设备单元又称为复合设备单元。蓝牙独特的组网方式赋予了它无线接入的强大生命力，同时可以有 7 个移动蓝牙用户通过一个网络节点与因特网相连。它靠跳频顺序识别每个微微网。同一微微网所有用户都与这个跳频顺序同步。

蓝牙分布式网络是自组网（Ad hoc Network）的一种特例。其最大特点是可以无基站支持，每个移动终端的地位是平等的，并可独立进行分组转发的决策。其建网灵活性，多跳性、拓扑结构动态变化和分布式控制等特点是构建蓝牙分布式网络的基础。

（3）标准发展现状

① IEEE 802.15.1

IEEE 获得蓝牙标准化团体 Bluetooth Special Interest Group（SIG）的授权，作为 IEEE Standard 802.15.1-2002 的基本部分可以对蓝牙标准的局部进行修改/复制。IEEE 批准的 IEEE 802.15.1 可以同 Bluetooth v1.1 完全兼容。

IEEE 802.15.1-2002 于 2002 年 6 月 14 日发布，对实现蓝牙设备的厂家来说，提供了所需要的补充标准。

② Bluetooth SIG

Bluetooth SIG（Bluetooth Special Interest Group，蓝牙技术联盟）是一家贸易协会，由电信、计算机、汽车制造、工业自动化和网络行业的领先厂商组成。该协会致力于推动蓝牙无线技术的发展，为短距离连接移动设备制定低成本的无线规范，并将其推向市场。

V1.1 为最早期版本，传输率约为 748～810kbit/s，因为是早期设计，容易受到同频率产品的干扰，影响通信质量。

V1.2 同样是只有 748～810kbit/s 的传输率，但加上了（改善 Software）抗干扰跳频功能。

V2.0 是 V1.2 的改良升级版，传输率为 1.8～2.1Mbit/s，可以有双工的工作方式，即在进行语音通信的同时，也可以传输档案、高像素图片。我国台湾地区有部分蓝牙 Dongle 已经在市面发售，但在手机内支持蓝牙 2.0 版本的情况则很少，蓝牙耳机能够真正使用的也不多，部分蓝牙产品自称是 2.0 版本，但仍然要利用外加配件才能达到预期效果。2.0 版本当然也支持 Stereo 运作。

V3.0 的颁布：2009 年 4 月 21 日，蓝牙技术联盟（Bluetooth SIG）正式颁布了新一代标准规范——Bluetooth Core Specification Version 3.0 High Speed（蓝牙核心规范 3.0 版高速）。

V4.0 是蓝牙技术联盟（Bluetooth Special Interest Group，SIG）在 2009 年 12 月正式发布的低功耗（low energy）版本蓝牙核心规格，仅需使用 1 颗标准钮扣电池，就可使用长达 1 年以上的时间。蓝牙将大幅扩展医疗保健、运动健身、安全管理、自动控制及家庭娱乐等市场。

3．WiFi

（1）技术介绍

WLAN（Wireless Local Area Network，无线局域网）是指应用无线通信技术将计算机设备互联起来，构成可以互相通信和实现资源共享的网络体系，从而使网络的构建和终端的

移动更加灵活。WLAN 通信系统一般使用在同一座建筑内。WLAN 使用 ISM（Industrial Scientific Medical）无线电广播频段通信。WLAN 的 IEEE 802.11a 标准使用 5GHz 频段，支持的最大速度为 54Mbit/s，而 IEEE 802.11b 和 IEEE 802.11g 标准使用 2.4GHz 频段，分别支持最大 11Mbit/s 和 54Mbit/s 的速度，IEEE 802.11n 使用 2.4GHz 和 5GHz 频段，支持最大 600Mbit/s 的速率。

目前正在研究的 IEEE 802.11ac 工作组目标速率在 1Gbit/s，工作在 5GHz 频段。

（2）基本参数

① 频带

频带位于 5GHz，2.4GHz。

不同标准的带宽分别为：

11a/b/g：20MHz；

11n：20MHz（必选），40MHz（可选）；

11ac：提升到 80/160MHz。

② 组网

WiFi 可以星形组网，在大型网络中支持 Mesh 组网。

在星形组网中，不同的 AP 芯片解决方案目前一般支持的最大接入用户数为 32。

③ 距离、速率

IEEE 802.11a 标准支持的最大速度为 54Mbit/s，IEEE 802.11b 和 IEEE 802.11g 标准分别支持最大 11Mbit/s 和 54Mbit/s 的速度，IEEE 802.11n 支持最大 600Mbit/s 的速率，802.11ac 工作组目标速率为 1Gbit/s。WiFi 的有效传输距离一般在 100 m 左右。有效传输距离与穿墙、信号干扰、多径等相关。

（3）标准发展现状

WLAN 标准的开发主要有两大组织机构：一个是 IEEE 的 802.11 工作组，一个是欧洲 ETSI 的 RES10 工作组。

① IEEE 802.11 工作组

第一个 802.11 标准是 1997 年完成的。它通过在 ISM 频段内使用扩频调制技术，提供高达 2Mbit/s 的数据传输速率。1999 年 9 月，IEEE 标准委员会又通过了 2 项最初标准的附录。第一个标准 802.11b，扩展了已存的 2.4GHz 物理层性能，使它的数据传输速率可以达到 11Mbit/s。第二个标准 802.11a，致力于在 5GHz 频段内的物理层中提供新的、更高的数据传输速率（20～54Mbit/s）。目前 IEEE 802.11ac 工作组已经成立，目标是研究 1Gbit/s 的数据传输速率。

② 欧洲 ETSI 的 RES10 工作组

除了 IEEE 802.11 家族，欧洲电信标准化协会（ETSI）的宽带无线电接入网络（BRAN）也制订出 HiperLAN 标准作为宽带无线接入网计划的组成部分，并在欧洲得到了广泛支持和应用。该系列包含 4 个标准：HiperLAN1、HiperLAN2、HiperLink 和 HiperAccess。HiperLAN1、HiperLAN2 用于高速 WLAN 接入；HiperLink 用于室内无线主干系统；HiperAccess 则用于室外对有线通信设施提供固定接入。

HiperLAN1 对应于 IEEE 802.11b，它工作在 5GHz，采用高斯滤波最小频移键控（GMSK）调制，速率上行可达 20Mbit/s。HiperLAN2 工作在 5GHz 频段，速率高达 54Mbit/s。

HiperLAN2 网络协议栈具有灵活的体系结构，很容易适配并扩展不同的固定网络。

③ WiFi 论坛组织

负责技术开发，市场培育以及严格的产品技术测试。WiFi 认证测试开始于 2000 年。它保证了参加测试的产品的互操作性以及质量，能够为最终用户提供最好的使用体验。目前测试通过的产品超过 5000 个。

7.8 终端演进策略建议

7.8.1 固网终端演进策略

IPv6 部署中，端、管、云三端有着天然的互动关系。电信运营中的 IT 支撑系统、业务网的 IPv6 化，也对固网终端有一定的要求。如 IMS 系统支撑的语音业务、IPTV 系统支撑的视频业务、ITMS 系统支撑的网管业务。

现网中的不同业务，分步演进，互不影响，如图 7-14 所示。现网中，接入终端通过不同的通道（PVC、VLAN）来承载不同业务。在双栈时代，业务平面可以独立在各自的 IPv4 或者 IPv6 网络中。由双栈迁移到 DS-Lite 后各业务执行情况如下。

① HSI 业务由双栈变成 IPv6 单栈，但涉及 IPv4 连接会跑在隧道中。

② IPTV 业务可保持 IPv4 流量不迁移，待系统支持 IPv6 后，升级 IPTV VLAN 到 IPv6。

③ VoIP 业务可保持 IPv4 流量不迁移，待部署支持 IPv6 的 SBC 后，升级 VoIP VLAN 到 IPv6。

④ ITMS 业务初期网管系统不改变现有 IPv4 网管体系，通过系统升级的方式实现对 IPv6 设备管理；后续升级改造现有网管系统管理功能，包括拓扑管理、故障管理、告警管理，支持 IPv6 标准 MIB 库和厂商的私有 MIB 库，实现对 IPv6 设备管理。

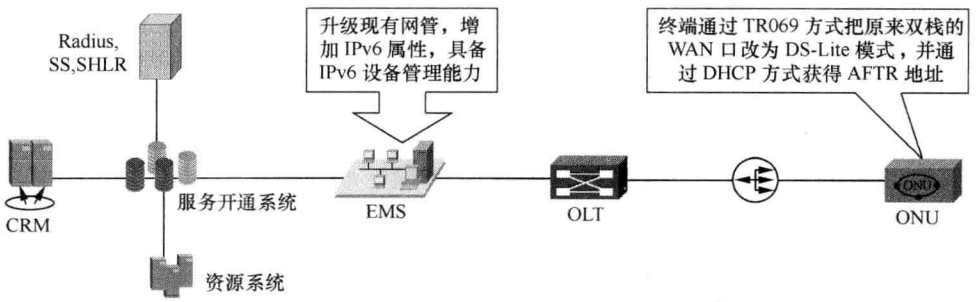


图 7-14 固网终端演进策略

对于运营支撑 IT 系统，为打造端到端无感知，对固网终端可以通过增加 IPv6 自动化工单，升级 CPE、ONU 等终端版本，增加 IPv6 属性，支持双栈、DS-Lite 标识业务等方法实现，如图 7-15 所示。

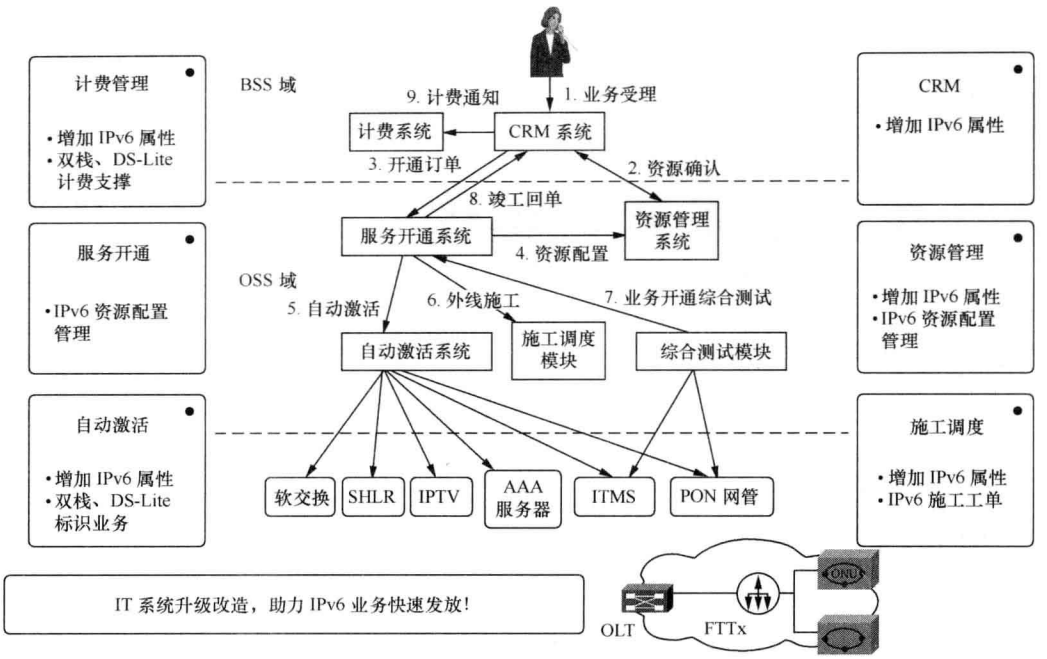


图 7-15　运营支撑 IT 系统改造方案

7.8.2　移动终端演进策略

目前，受 IPv6 应用推动不足及自身条件所限，移动终端对 IPv6 的支持现状较差，移动终端成为 IPv6 技术发展及推广的主要瓶颈之一。

为推动移动终端对 IPv6 的支持，需要在移动终端的基带芯片（Modem 解析方案）中实现对 IPv6 的支持。目前有少量终端已经完成相关开发工作，如，Motorola 在 2012 年 7 月推出的 Android 手机 XT889 已经宣称支持 IPv6 功能。同时，一些研究单位也提出了一种不依赖基带芯片支持 IPv6 的方案（Modem 透传方案），如中国电信北京研究院通过研究和实验，在 XT800、ME811 和 I909 等移动终端上完成了方案开发和实现，推动中国电信 3G 移动智能终端率先实现了 IPv6 技术。

尽管移动终端在技术上已经有了 IPv6 方案，但要实现向大规模商用的过渡，还需要很长一段路要走。最有效的办法是在移动终端入网时强制要求支持 IPv6 功能，扭转当前 IPv6 手机终端缺乏的局面。在手机的 IPv6 实现方式上，厂家的手机可以在依赖于芯片和不依赖于芯片两种方式中选择一种，但均以支持 IPv6 为基本要求。

当前中国的移动网络仍处在 3G 时代，在 IPv4 应用仍然占据主导地位的情况下，初期移动终端仍应以双栈方式为主，IPv4 地址不足的问题可考虑在网络侧通过地址复用技术来解决。在终端支持 IPv6 的同时，积极部署 IPv6 移动终端上的 IPv6 代表性应用和业务，推动 IPv6 移动终端的推广和普及，加速 IPv6 技术的发展。

未来几年内，第四代移动网络将在中国逐步部署实施。而 4G 移动网络和终端天然支持 IPv6，伴随着移动网络 4G 时代的来临，移动终端对 IPv6 的支持将向前迈进一大步。而且在 IPv6 应用普及的情况下，可考虑移动终端逐步关闭 IPv4 协议栈，以实现移动终端向下一代互联网的最终过渡。

第 8 章
IPv6 过渡技术部署实例

8.1 分布式 NAT444

8.1.1 实验内容与目标

① 试点场景：私网双栈＋分布式 NAT444＋PPPoE + 路由型 CPE。

② 试点业务：高速上网业务。

③ 试点用户：私网单/双栈用户、公网单/双栈用户、纯 IPv6 用户。

④ 试点网络：整个网络架构由用户接入 CPE、接入网、城域网、Internet 几个部分组成。城域网配有两台核心路由器，BRAS 采用 2*GE 方式双归属上行至城域出口核心路由器。每台 BRAS 上插入 2 块具备 NAT 功能的 CGN 单板互为主备。接入网主要采用两种接入方式，一种为 OLT + ONU 接入，一种为 DSLAM 接入。CPE 主要以路由型为例进行说明。

⑤ 试点内容：在城域网设备上配置双栈。在 BRAS 设备上配置不同的地址池并绑定到不同的域下。在 BRAS 上配置 NAT44 转换原则和方式，将用户绑定至分配的域名，BRAS 根据 AAA 下发的用户属性分配不同地址。BRAS 将为用户分配的“公网地址 + 端口块”上报 AAA。CGN 业务板卡间负载均衡、互为主备功能。

对于桥接性 CPE 用户，若用户终端支持双栈，则可以获取 IPv4 + IPv6 地址，若不支持，则仅获取 IPv4 地址；对于路由型 CPE 用户，若 CPE 不支持双栈，则用户依然仅获得 IPv4 地址。

8.1.2 网络结构

试点网络结构如图 8-1 所示。

如图 8-1 所示，用户/CPE 发起 PPPoE 连接，通过接入网终结于 BRAS，进而通过城域网接入 Internet。

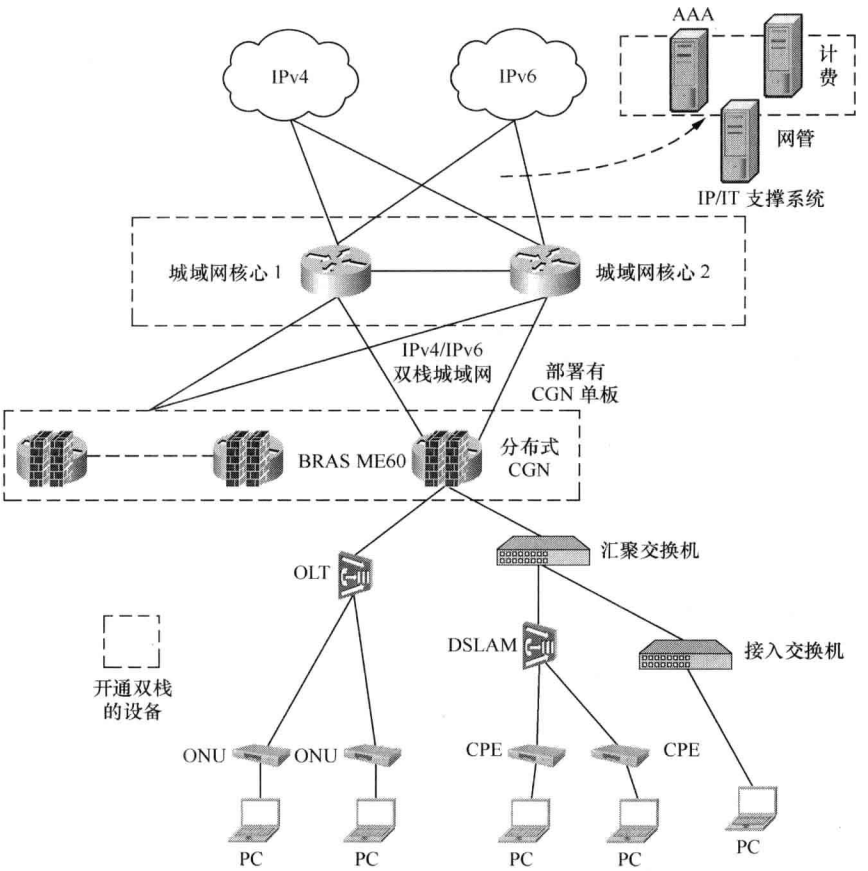


图 8-1　分布式 NAT444 试点网络结构示意图

8.1.3　业务流程

私网双栈＋分布式 NAT444＋PPPoE＋路由型 CPE 场景下用户上网流程如图 8-2 所示。

① CPE（路由模式）：CPE 与 BRAS 建立 PPPoE 连接。

② 在 PPPoE 之上，IPv4 和 IPv6 两个协议栈分别工作，获取 IPv4 私网地址和 IPv6 地址。

③ 用户 PC 获得 IPv4 私网地址后，使用 NAT444（分别在 CPE 和 CGN）转换为公网 IPv4 地址后，访问 IPv4 Internet。

④ 用户 PC 获得 IPv6 地址后，利用 IPv6 地址访问 IPv6 Internet。IPv6 直接在 IPv6 路由域转发，不经过 CGN。

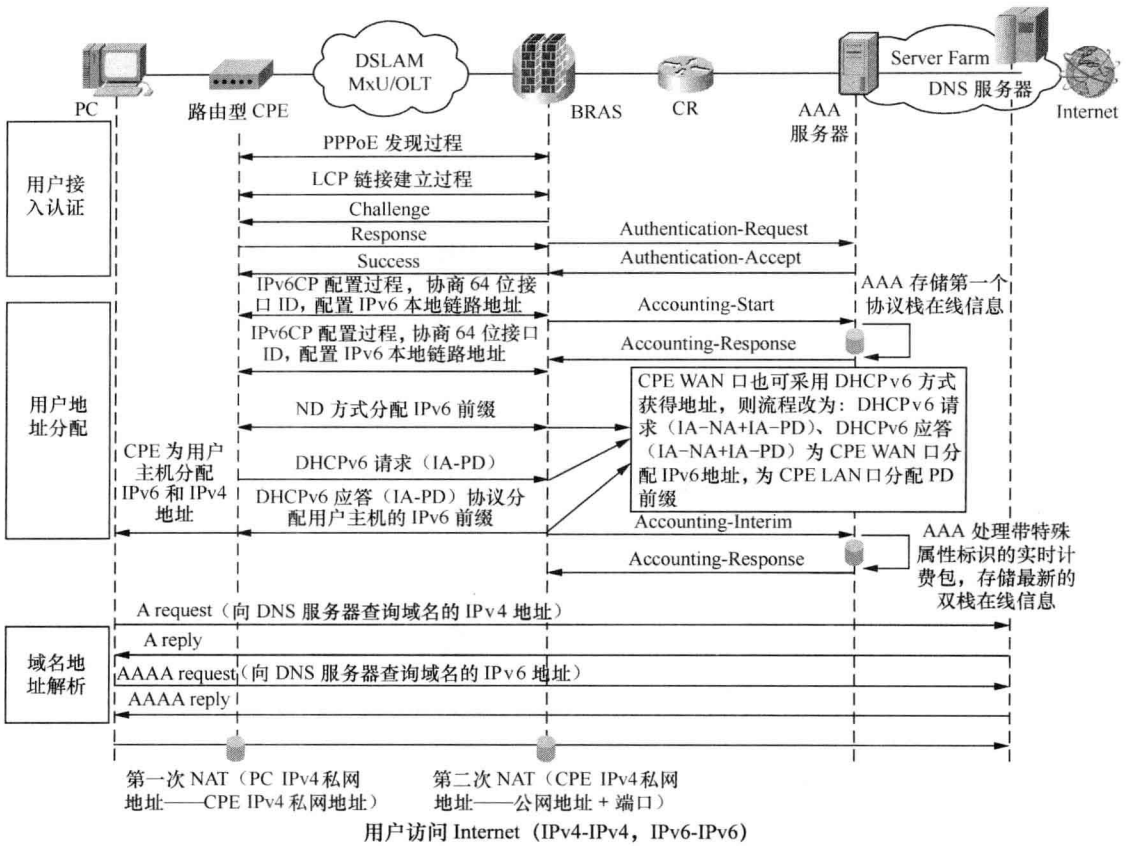


图 8-2　私网双栈 + 分布式 NAT444 + PPPoE + 路由型 CPE 场景下用户上网流程图

8.1.4　地址分配

对该场景 IPv6 地址的获取有 3 种方式，这 3 种方式中，路由型 CPE 的 LAN 口 PD 前缀均由 DHCP_IAPD 方式通过 DHCP 下发，但是 WAN 口地址的获取则有如下 3 种方式。

① 通过 ND_RA 获取。BRAS 与 CPE 进行 IPv6CP 协商交互，指定 CPE 的 WAN 口后/64 位 Interface-ID。BRAS 下发携带 IPv6 prefix 信息的 NDRA 报文，为每个 CPE 的 WAN 口独立分配 1 个/64 位的 IPv6 prefix 网段，该前缀和协商好的接口 ID 组成一个完整的地址，然后 CPE 再发起 DHCPv6 solicit 报文请求 PD 前缀。

② 通过 DHCP 方式获取。BRAS RA 通告中设置 M = 1，O = 1，CPE 发起 DHCPv6 solicit 报文，请求 WAN 口的 IA-NA 地址和 LAN 口的 PD 前缀，BRAS 应答，下发前缀。

③ 通过 unnumber 方式获取。这是一种简化方式，BRAS 中不需要再为 WAN 配置一个单独的地址池，直接从 PD 地址池中分配一个前缀即可。

上述几种方式均可采用，前两种用的比较多。

8.1.5　NAT 及其穿越

在私网双栈的场景下，用户获得私网 IPv4 地址后访问 Internet，需要在 CGN 处做 NAT，转换为公网地址。NAT 转换的方式有两种。一种是五元组模式，通过源地址、源端口号、协议类型、目的地址、目的端口号进行地址分配。这种方式安全，但是会造成相同的源地址和

源端口号被 NAT 转换为不同的外部地址和端口号，给地址溯源带来困难。另一种是依靠三元组，即源地址、源端口号、协议类型来转换。这种方式下相同的源和协议可以分配相同的公网地址和端口。只要建立了会话，公网上的主机也可以访问内网，可以支持广泛的 P2P 应用。实验中采用第 2 种方式，即三元组 NAT。

NAT 的方式也带来一个问题，那就是以 IP 地址和端口号作为通信标识符的应用层协议将无法工作，比如 SIP、通信协议等。这时需要采用一些 NAT 穿越技术（比如 ALG、STUN 等），或者对重要业务直接分配公网地址。实验中直接在 BRAS 上开启了 ALG 应用。

8.1.6　用户溯源

对于分布式 CGN 部署场景（即 BRAS 直接插 CGN 卡的部署方式），采用动态溯源方式，由 BRAS 生成每用户私网地址、公有地址、端口块的映射关系（保证每用户分到不同的公网地址和端口块），上报 AAA。

BRAS 需要通过端口预分配 Port-range 方式为用户地址选择公有地址及端口块。

8.1.7　设备具体配置

下面以华为设备为例，主要针对 BRAS（ME60）和 OLT 的配置加以说明。

BRAS 配置——IPv4 地址分配

[ME60]ip pool ipv4_ds_nat444 bas local //配置 IPv4 私网地址池给 CPE WAN 口分配私网地址。

[ME60-ip-pool-ipv4_ds_nat444]gateway 10.11.12.1 255.255.255.0

[ME60-ip-pool-ipv4_ds_nat444]section 1 10.11.12.2 10.11.12.254

[ME60-ip-pool-ipv4_ds_nat444]dns-server 172.16.2.6

[ME60]ip pool ipv4_public bas local //配置 IPv4 公网地址池给 CPE 分配公网地址。

[ME60-ip-pool-ipv4_public]gateway 136.10.2.1 255.255.255.0

[ME60-ip-pool-ipv4_public]section 1 136.10.2.2 136.10.2.254

[ME60-ip-pool-ipv4_public]dns-server 172.16.2.6

BRAS 配置——IPv6 地址分配

[ME60]ipv6 prefix oss_nd local //地址池属性配置为 local，为 CPE WAN 口分配的地址。

[ME60-ipv6-prefix-oss_nd]prefix 3001:10:: /48

[ME60]ipv6 prefix oss_pd delegation //地址池属性配置为 delegation，家庭网关为终端分配的地址。

[ME60-ipv6-prefix-oss_pd]prefix 3002:10::/48

[ME60]ipv6 pool oss_nd bas local //配置 IPv6 ND 地址池。

[ME60-ipv6-pool-oss_nd]prefix oss_nd //绑定 IPv6 ND 前缀到地址池。

[ME60]ipv6 pool oss_pd bas delegation //配置 IPv6 PD 地址池。

[ME60-ipv6-pool-oss_pd]dns-server 2001:1::E//配置 IPv6 DNS 服务器地址。

[ME60-ipv6-pool-oss_pd]prefix oss_pd

[ME60]DHCPv6 DUID LLT //配置设备的 DUID（DHCHv6 角色标志符）。

BRAS 配置——配置 NAT 实例

[ME60]user-group oss_ipv6_test_nat444

[ME60]nat session-table size 2M slot 1 //必须配置此命令，负责 NAT 转发，最大为 6M 支持 session。

[ME60]nat instance oss_ipv6-test //创建 NAT 实例。

[ME60-nat-instance-oss_ipv6_test]nat filter mode full-cone //配置 NAT。

[ME60-nat-instance-oss-ipv6_test]port-range 512 //配置用户 NAT 可用端口数，非必须配置，但是推荐根据需求配置。

[ME60-nat-instance-oss-ipv6_test]add slot 1 master //绑定 1 号 CGN 单板。

[ME60-nat-instance-oss-ipv6_test]address-group oss_ipv6_test_nat444 136.10.1.2 136.10.1.254 //配置 NAT 公网地址池。

[ME60-nat-instance-oss-ipv6_test]nat outbound any address-group oss_ipv6_test_nat444 //通过 ACL 规则确定哪些地址需要做进行地址转换，如不需要对地址进行限制，则选择“any”参数。

[ME60-nat-instance-oss-ipv6_test]nat alg all //开启 ALG 功能。

BRAS 配置——配置 Radius 服务器组

[ME60]radius-server group ds_nat444

[ME60-radius-ds_nat444]radius-server authentication 172.15.2.6 1812 weight 0

[ME60-radius-ds_nat444]radius-server accounting 172.15.2.6 1813 weight 0

[ME60-radius-ds_nat444]radius-server source interface LoopBack200 //指定和 Radius 服务器报文交互的源地址。

[ME60-radius-ds_nat444]radius-server share-key shiyan //密钥需要和服务器配置一致。

[ME60-radius-ds_nat444]quit

[ME60]radius-server authorization 172.15.6.2 share-key shiyan //配置授权服务器，用来下发属性。

BRAS 配置——策略配置

[ME60]aaa

[ME60-aaa]authentication-scheme nat444 //创建认证模板。

[ME60-aaa-authen-nat444]authentication-mode radius //配置认证为式为 Radius。

[ME60-aaa]accounting-scheme nat444 //配置计费模板。

[ME60-aaa-accounting-nat444]accounting-mode radius //配置计费方式为 Radius。

[ME60-aaa-accounting-nat444]quit

[ME60-aaa]domain ipv6_test //创建接入 DS + NAT444 用户域。

Info: Create a new domain.

[ME60-aaa-domain-ipv6_test]authentication-scheme nat444 //绑定认证模板。

[ME60-aaa-domain-ipv6_test]accounting-scheme nat444 //绑定计费模板。

[ME60-aaa-domain-ipv6_test]radius-server group ds_nat444 //进行 Radius 认证。

[ME60-aaa-domain-ipv6_test]ip-pool ipv4_ds_nat444 //绑定地址池。

[ME60-aaa-domain-ipv6_test]ipv6-pool oss_nd //绑定 IPv6 地址池。

[ME60-aaa-domain-ipv6_test]ipv6-pool oss_pd //绑定 IPv6 前缀池。

[ME60-aaa-domain-ipv6_test]ipv6 nd autoconfig managed-adderss-flag //配置 M 标识位值为 1，实现 IA_NA 方式分配地址（O 标识位默认值 1）。

[ME60-aaa-domain-ipv6-test]ipv6 nd autoconfig other-f1ag //配置 O 标识位值为 1。

[ME60-aaa-domain-ipv6_test]user-group oss_ipv6_test_nat444 bind nat instance oss_lpv6_test //user-group 绑定在域下，则该域下的用户都属于该 user-group，属于这个 user-group 的 IPv4 用户地址都要做 NAT 转换。

[ME60-aaa-domain-ipv6_test]accounting dual-stack identical //配置 IPv4 和 IPv6 统一计费。

[ME60-aaa]domain ipv6_public //创建公网地址用户域。

Info:Create a new domain

[ME60-aaa-domain-ipv6_public]authentication-scheme nat444 //绑定认证模板。

[ME60-aaa-domain-ipv6_public]accounting-scheme nat444 //绑定计费模板。

[ME60-aaa-domain-ipv6_public]radius-server group ds_nat444 //进行 Radius 认证。

[ME60-aaa-domain-ipv6_public]ip-pool ipv4_public //绑定公网地址池。

[ME60]interface Virtual-Template0

[ME60-Virtual-Template0]ppp authentication-mode chap //配置虚模极为 CHAP 认证。

BRAS 配置——接入业务的 BAS 子接口

[ME60]interface GigabitEthernet1/0/9.2005 //创建子接口。

[ME60-GigabitEthernet1/0/9.2005]pppoe-server bind Virtual-Template 0 //绑定 PPP 认证虚模板。

[ME60-GigabitEthernet1/0/9.2005]ipv6 enable //使能 IPv6 功能

[ME60-GigabitEthernet1l0/9.2005]ipv6 address auto link-local //这个配置主要作用是通过 ND 去协商客户端的下一跳地址，即 IPv6 默认路由。

[ME60-GigabitEthernet1/0/9.2005]user-vlan 2005 //终结用户 VLAN2005。

[ME60-GigabitEthernet1/0/9.2005]bas //使能 BAS 功能。

[ME60-GigabitEthernet1/0/9.2005-bas]access-type layer2-subscriber default-domain authentication ipv6_test //配置用户接入类型为二层接入用户，配置默认接入域名。

[ME60-GigabitEthernet1/0/9.2005-bas]arp-trigger //当 BAS 接口收到 ARP 报文就可以触发上线流程，用于用户异常下线重新认证的场景。

BRAS 配置——流量引入 CGN 单板

[ME60]acl number 6000 //创建用户访问控制列表（UCL），匹配用户组。

[ME60-acl-ucl-6000]rule 5 permit icmp source user-group oss_ipv6_test_nat444

[ME60-acl-ucl-6000]rule 10 permit ip source user-group oss_ipv6_test_nat444

[ME60]traffic classifier oss_ipv6_test //创建流分类。

[ME60-classifier-oss_ipv6_test]if-match acl 6000

[ME60]traffic behavior oss_ipv6_test //创建流行为，重定向流量到 NAT 板。

[ME60-behavior-oss_ipv6_test]nat bind instance oss_ipv6_test

[ME60]traffic policy oss_ipv6_test //创建流策略，把流分类和流行为关联起来。

[ME60-traffi cpolicy-oss_ipv6_test]classifier oss_ipv6_test behavior oss_ipv6_test

[ME60]traffic-policy oss_ipv6_test inbound //在系统下发到全局 inbound 方向，使符合条件的报文转发到 CGN 板上。

OLT 配置

创建业务 VLAN 并配置其上行口。

huawei>enable //进入到特权模式。

huawei#config //进入到配置模式。

huawei(config)#Vlan 2005 smat

huawei(config)#port vlan 2005 0/19 0 //业务 VLAN 为 2005，类型为 Smart VLAN，将上行端口 0/19/0 加入到 VLAN 2005 中。

配置 OBA（dynamic bandwidth allocation）模板，模板 ID 为 11，最大带宽为 2.048Mbit/s。

huawei(config)#dba-profile add profile-id 11 profile-name"ma5616"type4 max 2048

配置 GPON 线路模板。

huawei(config)#ont-lineprofile gpon profile-id 11 profile-name "ma5616" //增加 GPON ONT 线路模板，并进入 GPON ONT 线路模板模式。

huawei(config-gpon-lineprofile-11)#tcont 2 dba-profile-id 11 //使用 tcont 命令绑定下 CONT 与 OBA 模板。

huawei(config-gpon-lineprofile-11)#gem add 13 eth tcont 2 //增加索引为 13 的 GEM Port 用于承载管理业务流，GEM Port 13 绑定到 T-CONT 2。

huawei（config-gpon-lineprofile-11）#gem mapping 13 3 vlan 2005 //将管理业务流（用户侧 VLAN 为 2005）映射到索引为 13 的 GEM Port。

huawei(config-gpon-lineprofile-11)#commit //使用 commit 命令使模报配置参数生效。

配置 ONT 自动发现。

huawei(config)#interface gpon 0/14

huawei(config-if-gpon-0/14)#port 1 ont-auto-find enable

通过离线方式增加 ONT。

huawei(config–if-gpon-0/14)#ont add 1 1 sn-auth "4857544388507309" snmp ont-lineprofile-id 11 desc "ma5616"

创建业务流。

huawei(config)#service-port 2005 vlan 2005 gpon 0/14/1 ont 1 gemport 13 multi-service user-vlan 2005 tag-transforrn translate //创建连接 ONU 1 的业务流 2005 业务 VLAN 为 2005，GEM Port ID 为 13，用户侧 VLAN 为 2005。

8.2 集中式 NAT444

8.2.1 实验内容与目标

① 试点场景：私网双栈＋集中式 NAT444＋PPPoE＋路由型 CPE。

② 试点业务：高速上网业务。

③ 试点用户：私网单/双栈用户、公网单/双栈用户、纯 IPv6 用户。

④ 试点网络：整个网络架构由用户接入 CPE、接入网、城域网、Internet 几个部分组成。城域网配有两台核心路由器，BRAS 采用 2*GE 方式双归属上行至城域出口核心路由器。CGN 采用旁挂的方式双归属在城域网 CR 上。接入网主要采用两种接入方式，一种为 OLT + ONU 接入，另一种为 DSLAM 接入。CPE 以路由型为例进行重点说明。

⑤ 试点内容：在城域网设备上配置双栈。在 BRAS 设备上配置不同的地址池并绑定到不同的域下，将用户绑定至分配的域名，BRAS 根据 AAA 下发的用户属性分配不同地址。在 CGN 上配置相同的私网地址池和公网地址池，配置 NAT44 转换原则和方式。对于桥接型 CPE 用户，若用户终端支持双栈，则可以获取 IPv4 和 IPv6 地址，若不支持，则仅获取 IPv4 地址；对于路由型 CPE 用户，若 CPE 不支持双栈，用户依然仅获得 IPv4 地址。

8.2.2　网络结构

试点网络结构如图 8-3 所示。

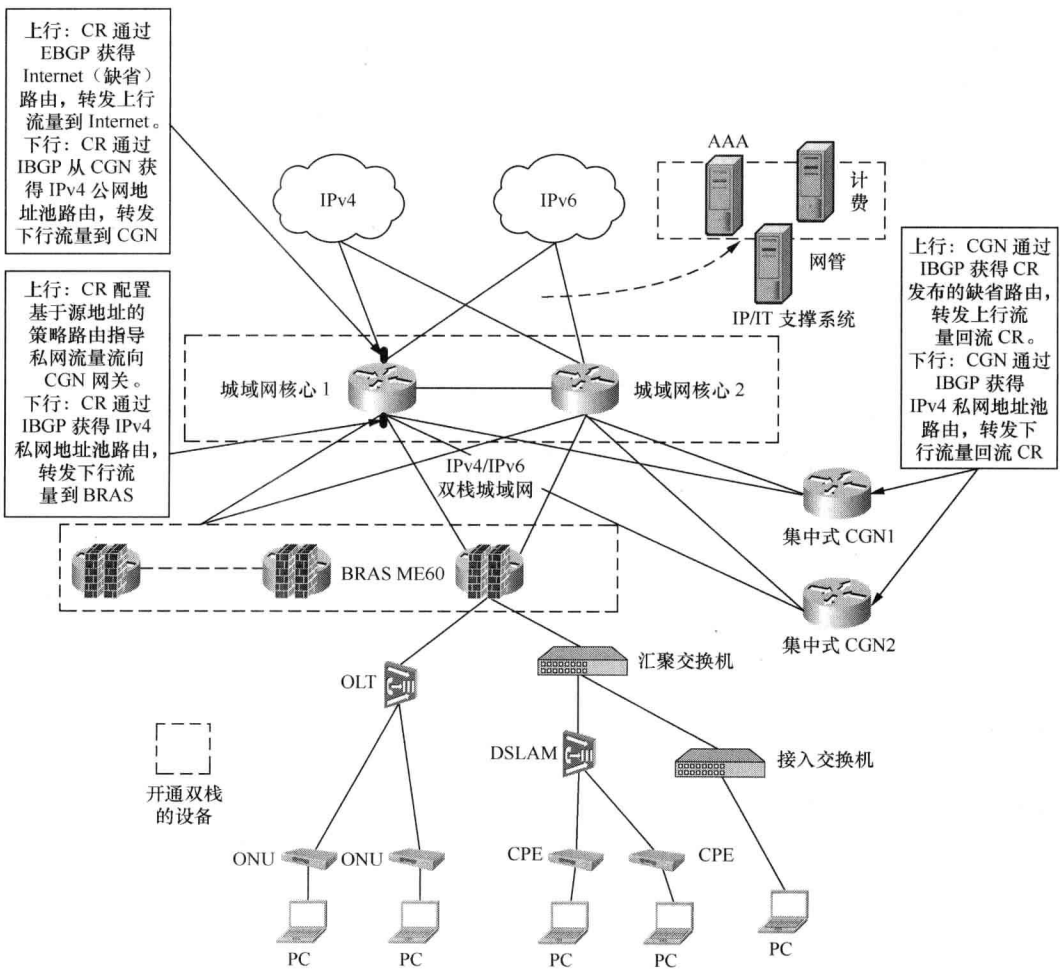


图 8-3　集中式 NAT444 试点网络结构图

如图 8-3 所示，用户/CPE 发起 PPPoE 连接，通过接入网终结于 BRAS，经过 BRAS 发给 CR 路由器。CR 路由器通过路由重定向策略将用户数据分组发送到集中式的 CGN 设备上，CGN 设备处理后再发给 CR 路由器，进而通过城域网接入 Internet。

8.2.3 业务流程

参见“8.1 分布式 NAT444”。

8.2.4 地址分配

地址分配的方式参见“8.1 分布式 NAT444”。

需要注意的是如果 CR 上配置路由策略，使得到达 CR1 的流量主选 CGN1，备选 CGN2（CR2 则相反）。由于 BRAS 双归 2 个 CR 并且是基于流做负载分担，同一个用户的不同流量可能被分担到不同 CR 上，进而引流到不同的 CGN，导致分配到不同的公网地址。这不符合规范且会导致溯源困难、某些应用无法访问（如 FTP、QQ、Telnet 等，采用源 IP 做 Session_ID，源 IP 变化导致 Session_ID 变化、Session 中断、业务中断）。

为了避免同一用户分配到不同的公网地址，必须修改路由策略，确保在 CGN 设备无故障的情况下，同一用户一定分配相同公网地址。具体做法为：在 CR 上配置强策略路由，使得来自 BRAS1 的流量优选 CGN1，来自 BRAS2 的流量优选 CGN2。

8.2.5 NAT 及其穿越

参见“8.1 分布式 NAT444”。

8.2.6 用户溯源

这种场景下，由于 NAT 转换在 CGN 上完成，BRAS 无法上报 AAA 公网地址，因此无法实现 Radius 溯源。这时可以采用 Syslog 溯源的方式。CGN 通过 Syslog 上报私网地址和公网地址 + 端口块映射关系，溯源服务器拼接从 SysLog 服务器和 AAA 获取的私网地址、用户信息、公网地址 + 端口块信息，并维持该映射关系。另外也可以采用静态算法溯源的方式。

8.2.7 设备配置

这种场景下，路由配置需要注意策略，转换的公网地址池配置在 CGN 上，其他配置参见“8.1 分布式 NAT444”。

8.3 分布式 DS-Lite

8.3.1 实验内容与目标

① 试点场景：分布式 DS-Lite（具备 NAT 功能）+PPPoE + 路由型 CPE（支持 DS-Lite）。
② 试点业务：高速上网业务。
③ 试点用户：纯 IPv6 用户、DS-Lite 用户。
④ 试点网络：整个网络架构由用户接入 CPE、接入网、城域网、Internet 几个部分组成。

城域网配有两台核心路由器，BRAS 采用 2*GE 方式双归属上行至城域出口核心路由器。每台 BRAS 上插入两块具备 NAT 功能的 AFTR 单板互为主备。接入网主要采用两种接入方式：一种为 OLT + ONU 接入，一种为 DSLAM 接入。CPE 以路由型为例进行重点说明，且需要具备 DS-Lite 隧道连接功能。

⑤ 试点内容：在 BRAS 到 AFTR 设备之间配置 IPv6 单栈即可，在 BRAS 到城域网核心设备之间需要配置双栈，相比前面全双栈的方案，城域网络是比较轻量的双栈网络。在 BRAS 设备上配置不同的地址池并绑定到不同的域下，在 BRAS 上配置 NAT44 转换原则和方式，将用户绑定至分配的域名，BRAS 根据 AAA 下发的用户属性分配不同地址。DS-Lite 用户只分配到 IPv6 地址。BRAS 将为用户分配的 IPv6 地址、公有地址、端口块等信息上报 AAA。AFTR 业务板卡间负载均衡、互为主备。

8.3.2　网络结构

试点网络结构如图 8-4 所示。

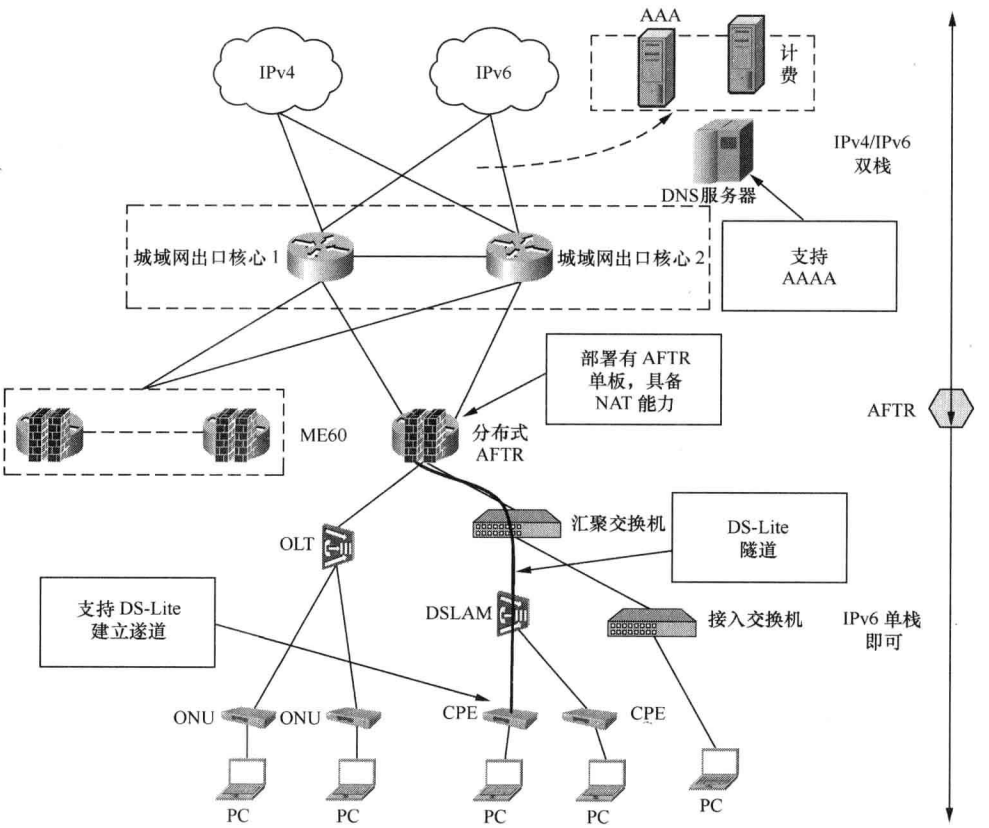


图 8-4　分布式 DS-Lite 试点网络结构图

如图 8-4 所示，CPE 发起 PPPoE 链接，通过接入网终结于 BRAS，进而通过城域网接入 Internet。因为 CPE 和 AFTR 之间是单栈 IPv6 网络，所以用户访问 IPv4 业务需要通过 CPE 和 AFTR 之间建立的 IPv4-in-IPv6 隧道进行传输。这条隧道终结于 AFTR。隧道的封装方式为 IPv4 报文直接封装 IPv6 的报文头。

8.3.3 业务流程

业务流程如图 8-5 所示。

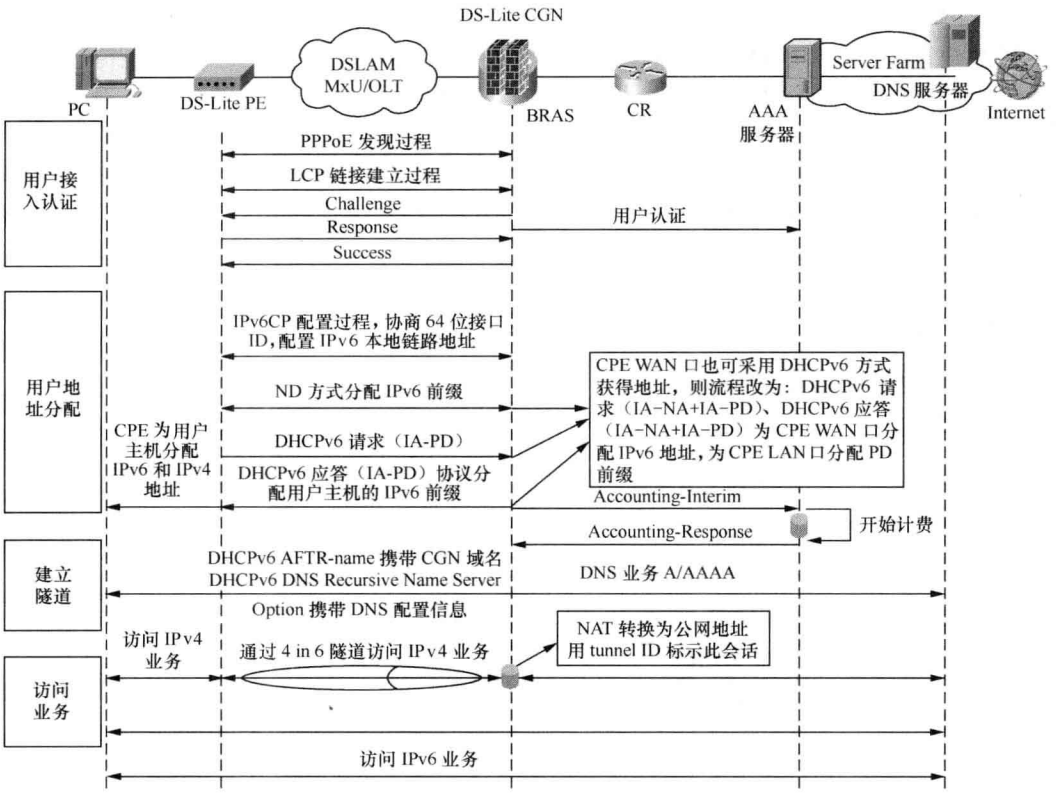


图 8-5 分布式 DS-Lite 业务流程图

① CPE 与 BRAS 之间发起 PPPoE 连接。

② 在 PPPoE 之上启动 IPv6 协议栈，BRAS 为 CPE 分配 IPv6 前缀和 DS-Lite AFTR 网关域名。

CPE 通过域名访问 DNS 获得 AFTR 网关地址。

③ CPE 通过 DHCP 方式为用户 PC 分配 IPv4 私网地址，通过 ND 或者 DHCPv6 方式为用户 PC 分配 IPv6 地址。

④ IPv6 报文通过 CPE、BRAS 直接转发。

⑤ IPv4 私网报文到达 CPE 后，被 CPE 封装在 IPv6 报文中，发给 DS-Lite AFTR 网关。

⑥ DS-Lite AFTR 对封装报文解封后作 NAT 44 后，转发至 IPv4 Internet。

8.3.4 地址分配

CPE 发起 DHCP RS 请求，BRAS 从本地地址池中选择 IPv6 前缀和地址，然后分配给 CPE。与双栈流程不同的是，BRAS 只给 CPE 分配 IPv6 地址。CPE 通过 DHCP 方式为用户 PC 分配 IPv4 私网地址，通过 ND 或者 DHCPv6 方式为用户 PC 分配 IPv6 地址。

DHCP 参数中会下发 AFTR 的域名。CPE 获得该域名后会主动发起 IPv6 DNS 请求，以获得对应的 IPv6 地址。访问 IPv4 的报文发至 CPE，触发 CPE 和 CGN 之间 4in6 隧道的建立。由此可见，IPv4 in IPv6 隧道是双向、无状态的隧道。

8.3.5　NAT 及其穿越

同私网双栈场景。

8.3.6　用户溯源

用户的私网 IPv4 地址由 CPE 分配，不能唯一标识一个会话，需要引入一些唯一的参数比如 tunnel ID、IPv6 等。

8.3.7　DNS 解析

DNS 需要支持双栈，当 DNS query 的是 AAAA 资源记录，DNS 服务器将域名解析为 IPv6 地址；当 DNS query 的是 A 资源记录，DNS 服务器将域名解析为 IPv4 地址。

对于 IPv6 DNS 业务，BRAS 通过 DHCPv6 协议将 IPv6 的 DNS 配置信息下发给 CPE。CPE 可以直接将 BRAS 下发的 DNSv6 地址下发给 PC，也可以将自己作为 DNS Server 将配置信息作为 DNS Proxy 下发给用户。

对于 IPv4 DNS 业务，可以让 CPE 实现 IPv4 DNS 和 IPv6 DNS 间的 Proxy 功能（只转换报文头）。或者为 CPE 配置静态的 DNS Server 的 IPv4 地址。

8.3.8　设备具体配置

下面以华为设备为例，主要针对 BRAS（ME60）加以说明。

配置 DS-Lite + PPPoE 场景个人 HSI 业务

配置 DS-Lite 实例

[BRAS]ds-lite instance dslite

[BRAS-ds-lite-instance-ds1ite]mode full-cone //配置 NAT 模式为三元组

[BRAS-ds-lite-instance-dslite]add slot 2 master //绑定 CGN 单板

[BRAS-ds-lite-instance-dslite]add slot 7 slave //配置 7 号 CGN 单板为备 NAT 板，可以实现 NAT session 的备份

[BRAS-ds-lite-instance-dslite]ds-lite address-group dslite 110.21.0.2 110.21 .0.101 //配置 DS-Lite 公网地址池。

[BRAS-ds-lite-instance-dslite]ds-lite outbound any address-group dslite //对 outbound 方向域下所有 IPv4 用户报文做地址转换。

[BRAS-ds-lite-instance-dslite]local-ip 2010:2222::0001 prefix-length 128 //指定本地 CGN 隧道地址，即 AFTR 地址。

[BRAS-ds-lite-instance-dslite]remote-ipv6 2011:201::prefix-length 64 //指定远端地址，该地址配置为 Local 地址池地址，否则 IPv4 业务不通。

IPv60 地址池内配置 AFTR 域名

[BRAS]ipv6 pool ds_cpe bas local
[BRAS-ipv6-pool-ds_cpe]aftr-name abc.fordemo.com

配置 DS-Lite 实例的安全特性

DS-Lite 实例下配置 port-range、port-imit、session-limit、log、trap、ALG 特性，命令请参考双栈 + NAT + PPPoE 中 NAT 实例的配置。

用户域下绑定 DS-Lite 实例。

[BRAS]user-group dsLite //创建 user-groupd。

[BRAS-aaa-domain-dslite]user-group dslite bind ds-lite instance dslite //在域下绑定 DS-Lite 实例。

说明：部署 DS-Lite4 需要 CPE 支持 AFTR 动态下发。

调测 DS-Lite + PPPoE 场景个人 HSI 业务

监测 DS-Lite tunnel 是否建立

[BRAS-2]display ds-lite tunnel table
Slot:7 Engine: 0
Current total sessions: 1
CPE:2011:0055::95cO:4578:bddd:8040 --> Local-ip:2121:0001:0001::0001
Slot: 7 Engine: 1
Current total sessions: 1
CPE:2011:0055::95c0:4578:bddd:8040 --> Local-ip:2121:0001:0001::0001
Slot: 7 Engine: 2
Current total sessions: 1.
CPE:2011.0055::95cO:4578:bddd:8040 --> Local-ip:2121:0001:0001::0001
Slot: 7 Engine: 3
Current total sessions: 0//查看 nat session 表

[BRAS-2]display nat session table slot 7

This operation will take a few minutes. Press 'Ctrl + C' to break ...
Slot: 7 Engine: 0
Current total sessions: 1.
icmp: 192.168.1.130:1[111.5.55.9:4096]-->1.1.0.14:2048
This operation will take a few minutes. Press 'Ctrl + C' to break ...
Slot: 7 Engine: 1
Current total sessions: 0.
This operation will take a few minutes. Press 'Ctrl + C' to break ...
Slot: 7 Engine: 2
Current total sessions' 0.
This operation will take a few minutes. Press 'Ctrl + C' to break ..
Slot: 7 Engine: 3
Current total sessions: 0

8.4　集中式 DS-Lite

8.4.1　实验内容与目标

① 试点场景：集中式 DS-Lite（具备 NAT 功能）＋PPPoE＋路由型 CPE（支持 DS-Lite）。

② 试点业务：高速上网业务。

③ 试点用户：纯 IPv6 用户、DS-Lite 用户。

④ 试点网络：整个网络架构由用户接入 CPE、接入网、城域网、Internet 几个部分组成。城域网配有 2 台核心路由器，BRAS 采用 2*GE 方式双归属上行至城域出口核心路由器。独立的 AFTR 设备采用旁挂方式双归属接入 CR。接入网主要采用两种接入方式，一种为 OLT＋ONU 接入，一种为 DSLAM 接入。CPE 以路由型为例进行重点说明，且需要具备 DS-Lite 隧道连接功能。

⑤ 试点内容：在 BRAS 设备上配置不同的地址池并绑定到不同的域下，将用户绑定至分配的域名，BRAS 根据 AAA 下发的用户属性分配不同地址。DS-Lite 用户只分配到 IPv6 地址。在 AFTR 上配置 NAT44 转换原则和方式。

8.4.2　网络结构

网络结构如图 8-6 所示。

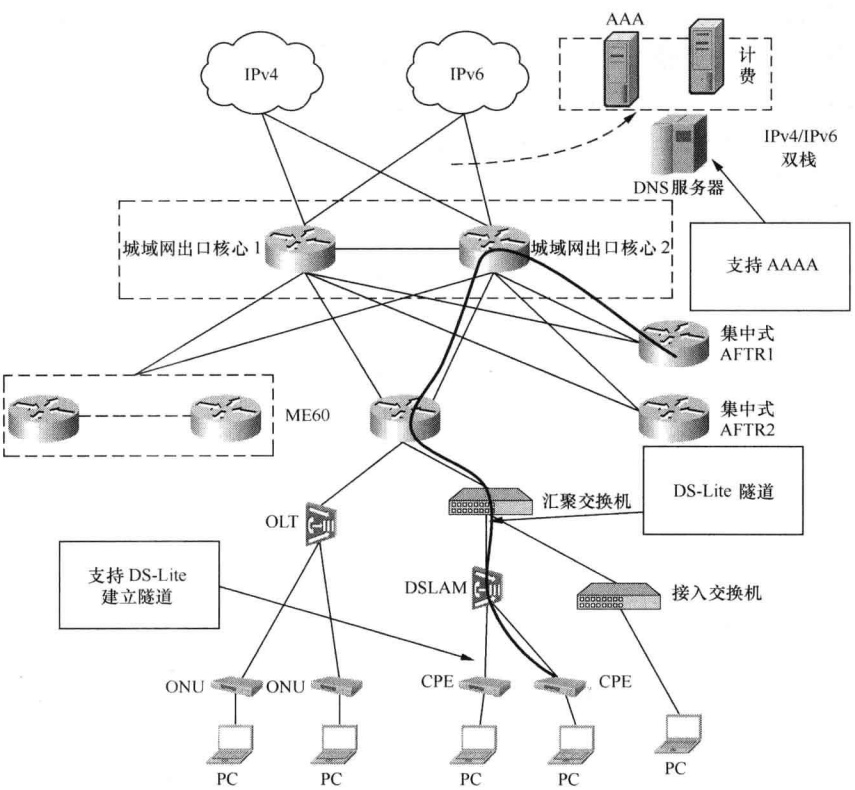


图 8-6　集中式 DS-Lite 试点网络结构图

如图 8-6 所示，CPE 发起 PPPoE 连接，通过接入网终结于 BRAS，进而通过城域网接入 Internet。因为 CPE 和 AFTR 之间是单栈 IPv6 网络，所以用户访问 IPv4 业务需要通过 CPE 和 AFTR 之间建立的 IPv4-in-IPv6 隧道进行传输，这条隧道终结于 AFTR。隧道的封装方式为 IPv4 报文直接封装 IPv6 的报文头。

8.4.3 其他

参见 8.2 节和 8.3 节。

8.5 IPv6 过渡技术实验小结

双栈和 DS-Lite 两种方案的比较见表 8-1。

表 8-1　　双栈和 DS-Lite 方案对比表

方案	优点	缺点	使用场景
双栈	IPv4、IPv6 业务独立，后期向 IPv6 演进的过程中可逐渐关闭 IPv4； 双栈方案和 NAT 技术成熟，应用较多	双栈管理相对复杂，占用系统资源； 要解决 IPv4 地址不足的问题必须配置私网 NAT	IPv4 和 IPv6 业务将在一段时间内长期同时共存。适合已有的用户
DS-Lite	BRAS 只需要 IPv6 单栈，用户管理简单； 容易迁移到纯 IPv6 网络	CPE 需要支持 DS-Lite； IPv4 业务通过 IPv6 隧道承载不容易分辨故障	CPE 设备需要支持 DS-Lite，适合新建网络的用户

由此可见，两种方案技术上都比较成熟，各有优势，适用于不同的场景。

实验也暴露出 IPv6 及其过渡技术的引入带来了一系列需要解决的问题，如，私网地址溯源困难；某些上层应用无法穿越；IPv6 资源少，用户体验不佳，等等。因此商用部署还需要各个层面进一步的推动。

第 9 章
IPTV 业务 IPv6 过渡部署实例

9.1 实验概况

9.1.1 IPTV 现有业务系统介绍

本章以中兴通讯建设的现有 IPTV 业务系统为例介绍。

（1）系统架构

IPTV 系统在总体结构上分为 5 层，即内容运营系统、业务运营系统、CDN 分发网络、承载网络和家庭网络。系统架构如图 9-1 所示。

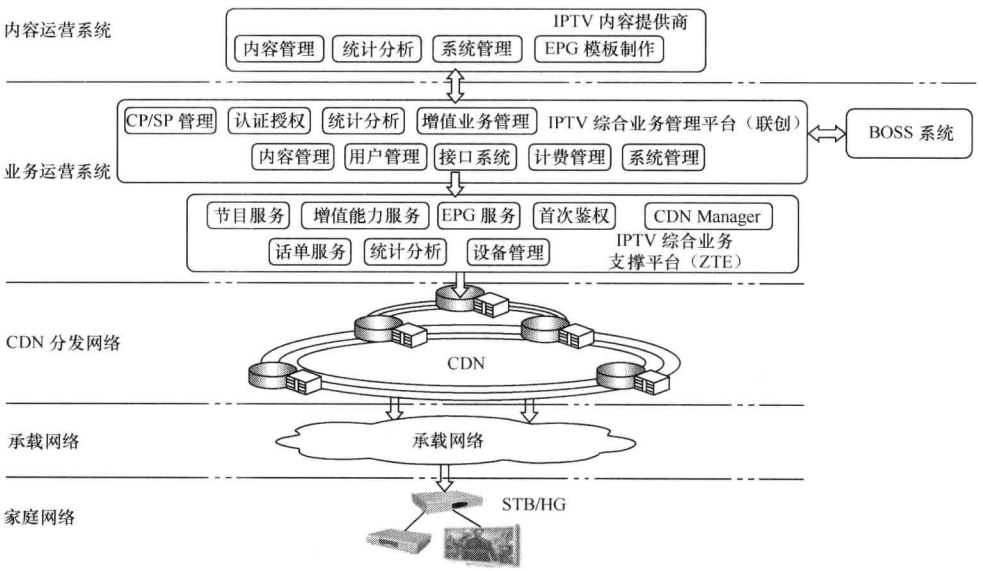


图 9-1 IPTV 系统架构图

（2）EPG（Electronic Programmer Guide）

电子节目单，向用户提供电视节目菜单展示。

（3）CDN（Content Delivery Network）

内容分发服务网络 CDN 分发网络完成视频数据的导入、存储、分发和服务等功能，设

备主要包括流媒体服务器和存储设备。本系统中的 CDN 分发网络分为 3 层架构，即：省中心、VOD/TVOD 区域中心和边缘节点 3 层架构。由边缘节点为用户直接提供视频流服务。系统可以提供的业务包括电视广播、视频点播、时移电视、增值业务等。

IPTV 业务系统承载在已有的城域网和接入网之上，CDN 的区域和边缘节点与城域网出口核心相连，用户接入部署支持 IPTV 和宽带的家庭网关。IPTV 直播流量通过专网由城域网接入分发，具体网络结构如图 9-2 所示。

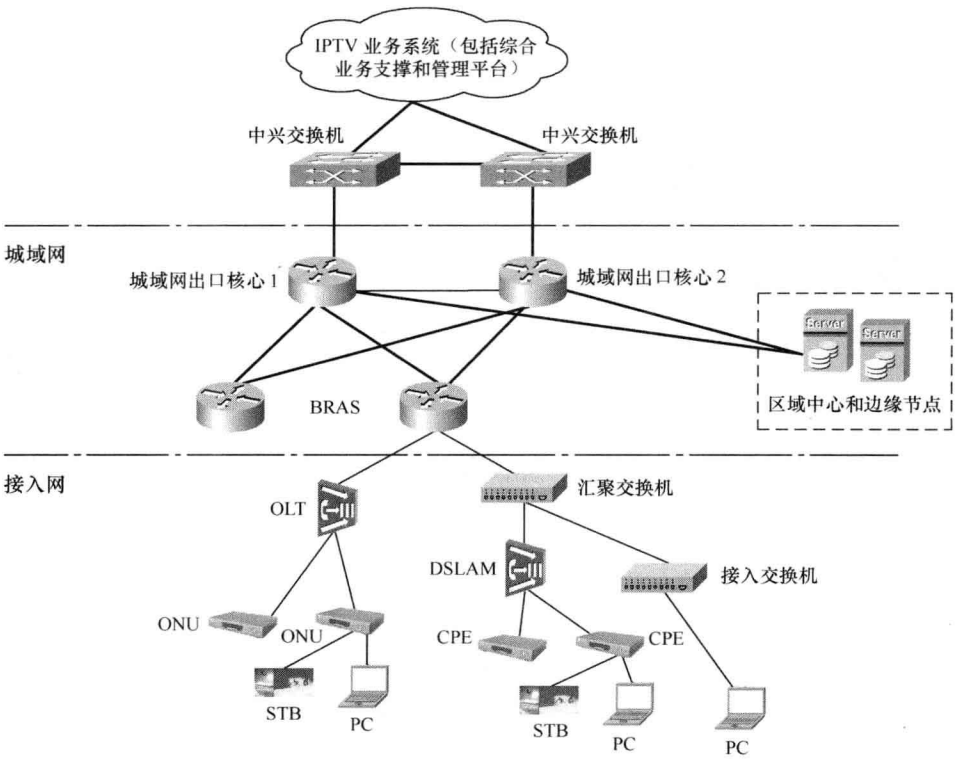


图 9-2　IPTV 承载网络结构图

上述已有系统完全是基于 IPv4 构建的。

9.1.2　实验目标

① 基于现有已具备开通双栈能力的城域网，建设 IPv6 试验网络，实现 IPTV 所有业务及增值业务在 IPv6 环境下完成端到端业务流程。为启动现网 IPTV 业务 IPv6 改造积累实际经验。

② 参考各种 IPTV 接入场景，验证各主流厂家现网设备对 IPv6，尤其是 IPv6 多播业务的支持能力。

9.2　过渡部署方案及流程

9.2.1　现有系统改造方案

IPTV 业务 IPv6 改造涉及的所有网元和具体改造内容整体的技术分析如图 9-3 所示。

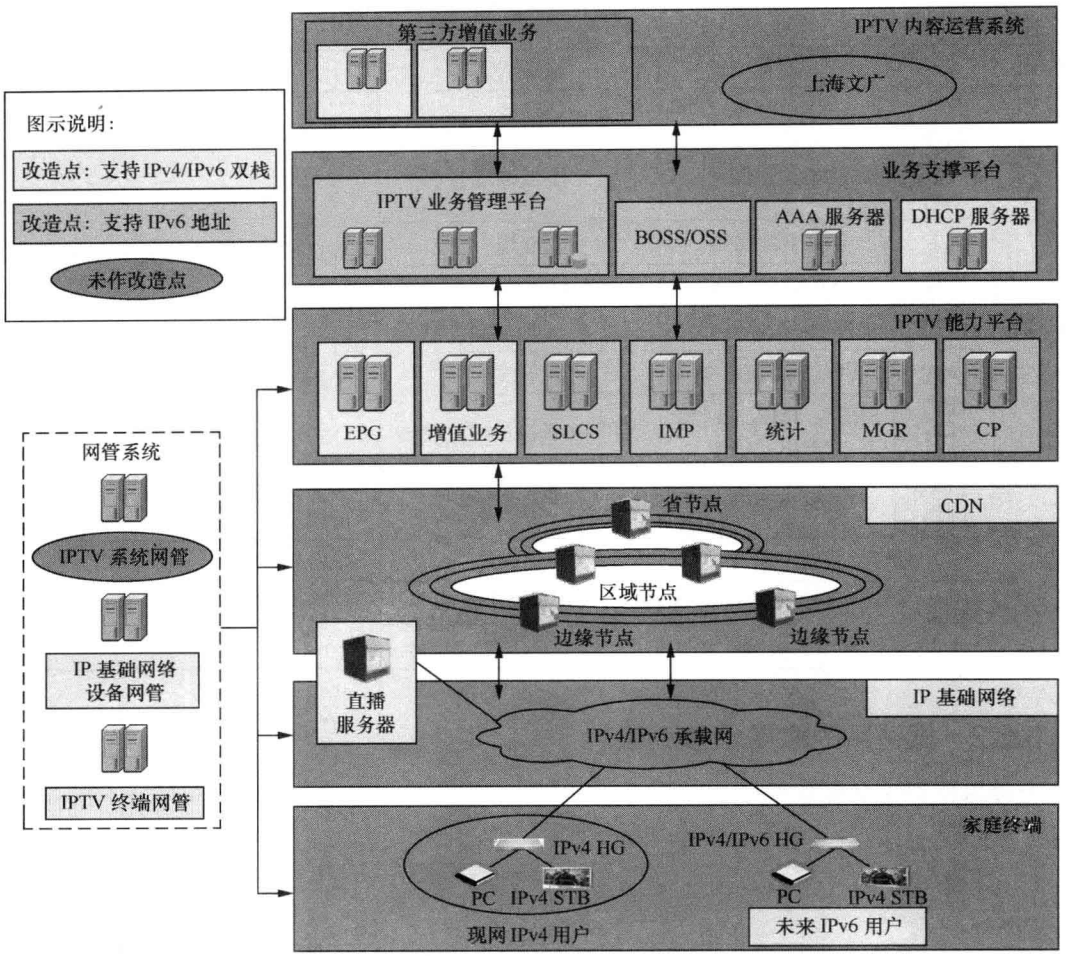


图 9-3　改造分析图

由图 9-3 可见，承载网络和直播服务器必须支持双栈，与访问流对接的 EPG 系统也必须支持双栈。

图中 IPTV 能力平台为中兴公司平台内部设备，即：SLCS 为业务及管理服务器，CP 为内容处理，MGR 是管理服务器，IMP 是业务接口机。

各网元的改造内容及最终达成目标见表 9-1。

表 9-1　　网元改造内容及最终目标

系统		设备	改造内容	达成目标
用户	无	无	无	改造过程是透明的，用户对于现有电信业务的体验无任何下降
终端	终端	家庭网关	双栈	上网业务：支持 IPv4/IPv6 双栈访问
		STB 机顶盒	支持双栈，但一般启用单栈模式（IPv4 或 IPv6）	IPTV 业务：支持 IPv4 或者 IPv6 终端访问
IP 基础网络	网络	接入网	双栈	实现 IPTV、宽带等多业务承载，确保现有电信业务不受影响
		SR/BRAS	双栈	

续表

系统		设备	改造内容	达成目标
IPTV 系统侧	IPTV 内容运营系统	增值业务	双栈	同时满足现网 IPv4 和未来 IPv6 机顶盒用户接入。 确保现网 IPv4 用户享受服务不受影响，IPv6 新增用户的用户体验不低于 IPv4 用户
		广电其他网元	不做改造	
	业务支撑平台	DHCP 服务器	双栈	
		AAA、BOSS	支持 IPv6 地址属性管理	
		IPTV 业务管理平台	支持 IPv6 地址属性管理	
	IPTV 能力平台	EPG、增值业务	双栈	
		SLCS、IMP、统计、MGR、CP	支持 IPv6 地址属性管理	
	CDN 网络	CDN 网络节点	双栈	
网管系统	IPTV 系统网管		不做改造	同时实现 IPv4/IPv6 设备的网络管理
	IPTV 终端网管		双栈	
	IP 基础网络设备网管		双栈	

9.2.2 改造后的点播业务流程

业务流程如图 9-4 所示。

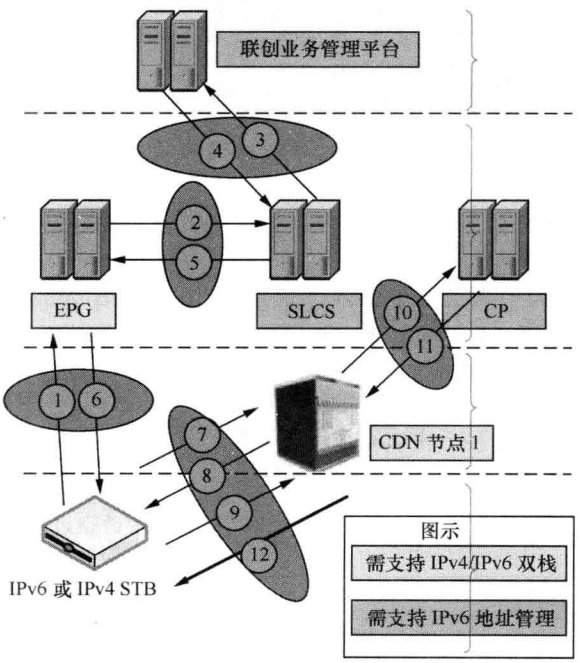


图 9-4　点播业务流程图

通信过程见表 9-2。

表 9-2　　点播业务通信过程

步骤	业务流程	协议类型	接口类型
1	用户通过 IPv4 或 IPv6 STB 上发起点播请求	IPv4 或 IPv6	内部接口
2	EPG 向业务系统 SLCS 发起鉴权操作	IPv4	内部接口
3	SLCS 发现该业务请求需要向联创业务管理平台进行鉴权，则请求向联创业务管理平台鉴权（可选流程）	IPv4	外部接口
4	联创业务管理平台业务鉴权通过（可选流程）	IPv4	外部接口
5	业务系统 SLCS 通过处理，返回 EPG 鉴权结果	IPv4	内部接口
6	EPG 向 STB 返回鉴权通过消息，并携带归属 CDN 节点的 URL 信息	IPv4/IPv6	内部接口
7	STB 向 CDN 节点 1 请求内容 URL	IPv4/IPv6	内部接口
8	CDN 节点 1 向 STB 返回可以提供服务的消息	IPv4/IPv6	内部接口
9	STB 根据上步的结果请求媒体服务	IPv4/IPv6	内部接口
10	CDN 节点 1 通知 CP 服务开始	IPv4	内部接口
11	CP 记录服务开始并返回给 CDN 节点 1	IPv4	内部接口
12	CDN 向用户提供媒体流服务	IPv4/IPv6	内部接口

9.2.3　改造后的直播业务流程

改造后直播业务流程如图 9-5 所示。

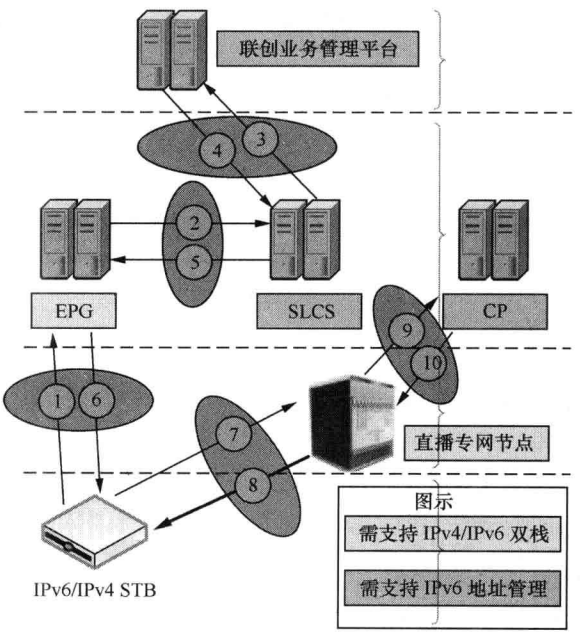


图 9-5　直播业务流程图

通信过程见表 9-3。

表 9-3　　直播业务通信过程

步骤	业务流程	协议类型	接口类型
1～6	同点播流程	同点播流程	同点播流程
7	STB 请求直播专网节点提供直播服务	IPv4/IPv6	内部节点
8	直播专网节点给 STB 发送直播码流	IPv4/IPv6	内部节点
9	直播专网节点向 CP 发送“服务开始”请求	IPv4	内部节点
10	开始服务确认	IPv4	内部节点

9.2.4　改造后的增值业务流程

业务流程如图 9-6 所示。

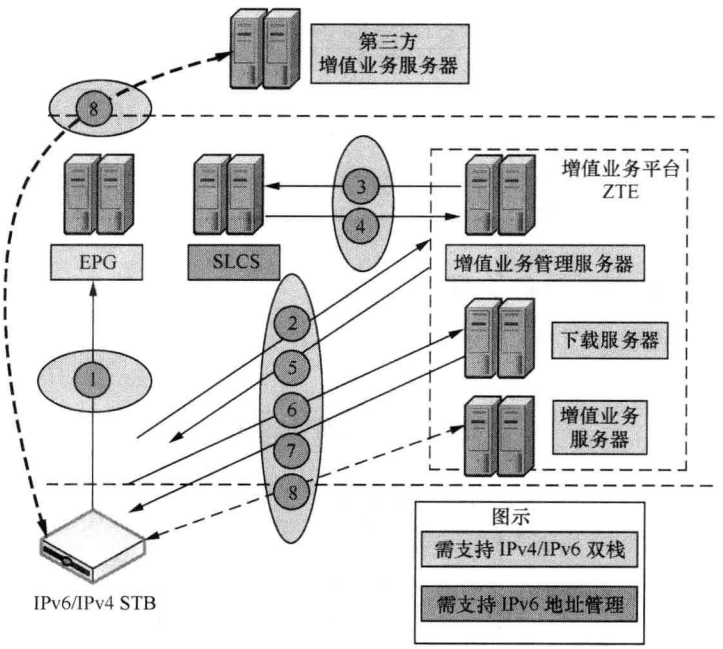


图 9-6　增值业务流程图

具体分为以下 3 种情况。

① 电信封闭的增值业务（图 9-6 中虚线框内所示）：双栈改造，满足 IPv4/IPv6 用户需求。

② 第三方增值业务（图 9-6 中虚线箭头所示，不包含媒体业务）：建议第三方进行双栈改造。若增值业务无法升级，现阶段考虑在 IP 基础网络部署 NAT 来解决。考虑到增值业务本身流量不大，IP 基础网络 NAT 可集中部署在 IDC 机房出口处。但某些业务服务器可能是第三方自行部署的，对于这样的情况，建议第三方 SP 逐步迁到 IDC 机房内。

③ 第三方增值业务（媒体业务）：建议第三方进行双栈改造，否则 IPv4 终端无法访问该业务。

通信过程见表 9-4。

表 9-4　　增值业务通信过程

步骤	业务流程	协议	接口类型
1	用户访问 EPG，点击增值业务应用链接	IPv4/IPv6	内部接口
2	用户访问增值业务应用平台，选择使用增值业务	IPv4/IPv6	内部接口
3	增值业务应用平台向 SLCS 发起增值业务的鉴权	IPv4	内部接口
4	SLCS 向增值业务应用平台返回鉴权结果	IPv4	内部接口
5	增值业务应用平台向 STB 游戏客户端的下载 URL 信息	IPv4/IPv6	内部接口
6	STB 向下载服务器请求下载游戏客户端	IPv4/IPv6	内部接口
7	STB 从下载服务器下载游戏客户端	IPv4/IPv6	内部接口
8	STB 启动客户端与游戏服务器通信	IPv4/IPv6	内部接口或外部接口

9.2.5　改造后的上网业务流程

业务流程如图 9-7 所示。

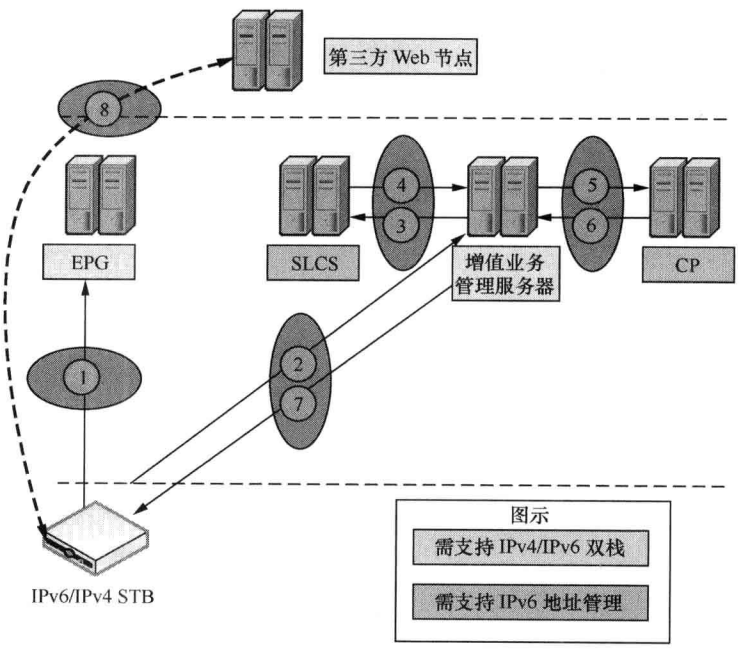


图 9-7　上网业务流程图

通信过程见表 9-5。

表 9-5　　上网业务通信过程

步骤	业务流程	协议类型	接口类型
1	用户访问 EPG，点击增值业务应用链接	IPv4/IPv6	内部接口
2	用户访问增值业务应用平台，选择使用上网增值业务	IPv4/IPv6	内部接口
3	增值业务应用平台向 SLCS 发起增值业务的鉴权	IPv4	内部接口

续表

步骤	业务流程	协议类型	接口类型
4	SLCS 向增值业务应用平台返回鉴权结果	IPv4	内部接口
5	增值业务应用平台通知 CP 计费开始	IPv4	内部接口
6	CP 记录服务开始并返回给增值业务平台	IPv4	内部接口
7	增值业务平台将要访问的站点 URL 信息返回给 STB	IPv4/IPv6	内部接口
8	STB 访问返回的站点	IPv4/IPv6	外部接口

在 STB 启用上网业务，访问的网站将是受控的。建议运营商开放上网业务时，确保 Web 应用完成双栈改造。

缩略语

缩略语	英文全称	中文全称
1x	cdma2000 1x 缩写	3G 移动通信标准
3GPP	The 3rd Generation Partnership Project	第三代合作伙伴计划
AAA	Authentication；Authorization；Accounting	认证；授权；计账
ABR	Area Border Router	区域边界路由器
ACL	Access Control List	访问控制列表
ADSL	Asymmetric Digital Subscriber Line	非对称数字用户线
AFTR	DS-Lite Address Family Transition Router element	DS-Lite 地址翻译网元
ALG	Application Level Gateway	应用层网关
ASBR	Autonomous System Border Router	自治系统边界路由器
BRAS	Broadband Remote Access Server	宽带远程接入服务器
BSR	Bit Scan Reverse	比特扫描扭转
C-BSR	Candidate-BSR	自举荐路由器
cdma2000	Code Division Multiple Access 2000	码分多址 2000，一种 3G 移动通信标准
CDN	Content Delivery Network	内容分发网络
CE	Consumer Edge router	用户边缘路由器
CGN	Carrier Grade NAT	运营级 NAT 设备
CHAP	Challenge Handshake Authentication Protocol	询问握手认证协议
CNAME	Canonical Name	别名记录
CNGI	China's Next Generation Internet	中国下一代互联网
CP/SP	Content Provider/Service Provider	内容/服务提供商
CPE	Customer Premises Equipment	用户驻地设备
CR	Core Router	核心路由器
CRLF	Carriage-Return Line-Feed	回车换行
C-S	Client-Server	客户端—服务器模式
DB	DataBase	数据库
DCCI	Data Center of CHINA Internet	中国互联网数据中心
DHCP	Dynamic Host Configuration Protocol	动态主机配置协议

缩略语	英文全称	中文全称
DNS	Domain Name Server	域名服务器
DR	Designated Router	指定路由器
DSLAM	Digital Subscriber Line Access Multiplexer	数字用户线路接入复用器
DS-Lite	Dual-Stack Lite	轻量级双栈
D-V	Distance Vector Routing Algorithm	距离矢量路由算法
EPG	Electronic Programmer Guide	电子节目单
FTP	File Transfer Protocol	文件传输协议
GE	Gigabit Ethernet	吉比特以太网
GGSN	Gateway GPRS Support Node	网关 GPRS 支持节点
GRT	Graft Retry Timer	定时器
GUA	Global Unicast Address	全局单播地址
HTTP	HyperText Transport Protocol	超文本传送协议
IANA	Internet Assigned Numbers Authority	因特网数字分配机构
IA-NA	Identity Association-Network Address	地址集合
IAPD	Identity Association-Prefix Delegation	前缀集合
ICMP	Internet Control Message Protocol	因特网控制消息协议
ICMPv6	Internet Control Message Protocol Version 6	第六版因特网控制消息协议
ICP	Internet Content Provider	因特网内容提供者
IDRP	Inter Domain Routing Protocol	域间路由选择协议
IGMP	Internet Group Management Protocol	Internet 组管理协议
IGP	Interior Gateway Protocol	内部网关协议
IMAP	Internet Mail Access Protocol	交互式邮件存取协议
IP	Internet Protocol	因特网协议
IPSec	Internet Protocol Security	Internet 协议安全性
IPX	Internet Packet Exchange protocol	因特网分组交换协议
IS-IS	Intermediate System to Intermediate System	中间系统到中间系统
ISP	Internet Service Provider	因特网服务提供者
ITU	International Telecommunications Union	国际电信联盟
L2TP	Layer 2 Tunneling Protocol	二层隧道协议
LACP	Link Aggregation Control Protocol	链路聚合控制协议
LAN	Local Area Network	局域网
LCP	Link Control Protocol	链路控制协议
LSA	Link-State Advertisement	链路状态广播
LTE	Long Term Evolution	长期演进
MID	Management Information Datsbase	管理信息库
MLD	Multicast Listener Discover	多播侦听发现协议
MMS	Microsoft Media Server protocol	微软的流媒体服务器协议
MPLS	Multi-Protocol Label Switching	多协议标签交换

缩略语	英文全称	中文全称
MSDP	Multicast Source Discovery Protocol	多播源发现协议
MTU	Maximum Transmission Unit	最大传输单元
NAT	Network Address Translation	网络地址转换
NAT64	IPv6-IPv4 Network Address Translation	IPv6-IPv4 网络地址转换
NAT-PT	Network Address Port Translation	网络地址端口转换
NBMA	Non-Broadcast Multiple Access	非广播多路访问网络
NGN	Next Generation Network	下一代网络
NSAP	Network Service Access Point	网络服务接入点
OLT	Optical Line Terminal	光线路终端
ONU	Optical Network Unit	光网络单元
OSPF	Open Shortest Path First	开放式最短路径优先
P2P	Peer to Peer	对等网络
P4P	Proactive network Provider Participation for P2P	P2P 协议升级版
PAP	Password Authentication Protocol	密码认证协议
PD	Prefix Delegation	前缀集合
PDA	Personal Digital Assistant	掌上电脑
PDSN	Packet Data Serving Node	分组数据服务节点
PE	Provider Edge router	运营商边缘路由器
PIM	Protocol Independent Multicast	协议无关多播
PMTU	Path MTU	路径 MTU
PNM	Progressive Networks Audio	Real 专用的实时传输协议
POP3	Post Office Protocol 3	邮局协议的第 3 个版本
PPPoE	Point-to-Point Protocol over Ethernet	在以太网上承载点到点协议
QoS	Quality of Service	服务质量
RAN	Radio Access Network	无线接入网
RIB	Resource Information Block	资源信息块
RIP	Routing Information Protocol	路由信息协议
RIPNG	RIP for IPv6	下一代路由选择信息协议
RIR	Regional Internet Register	地区性 Internet 注册机构
RP	Rendezvous Point	汇聚点
RPT	RP-rooted Shared Tree	基于 RP 的共享树
RSVP	Resource Reservation Protocol	资源预留协议
RTCP	Real-time Transport Control Protocol	实时传输控制协议
RTP	Real time Transport Protocol	实时传输协议
RTSP	Real Time Streaming Protocol	实时流媒体协议
SIP	Session Initiation Protocol	会话初始协议
SMTP	Simple Mail Transfer Protocol	简单邮件传输协议
SPF	Shortest Path First	最短路径优先算法

缩略语	英文全称	中文全称
SPI	Serial Peripheral Interface	串行外设接口
SPT	Shortest Path Tree	最短路径树
SSL	encrypted Secure Sockets Layer	安全套接层，网络安全协议
SSM	Source Specific Multicast	源特定多播
STUN	Simple Traversal of UDP over NATs	NAT 的 UDP 简单穿越
TCP	Transmission Control Protocol	传输控制协议
TD-SCDMA	Time Division-Synchronous Code Division Multiple Access	时分同步码分多址
TLS	Transport Layer Security	安全传输层协议
UDP	User Datagram Protocol	用户数据报协议
UIM	User Identity Module	用户识别模块
URI	Uniform Resource Identifier	通用资源标识符
URL	Universal Resource Locator	统一资源定位符
VLL	Virtual Leased Line	虚拟租用线
VolP	Voice over IP	利用 IP 网络实现语音通信的一种先进通信手段
VPLS	Virtual Private LAN Service	虚拟专用局域网业务
VPN	Virtual Private Network	虚拟专用网
VRRP	Virtual Router Redundancy Protocol	虚拟路由器冗余协议
WAN	Wide Area Network	广域网
WCDMA	Wideband Code Division Multiple Access	宽带码分多址
XPON	X-Passive Optical Network	新一代光纤接入技术