



ciscopress.com



TCP/IP路由技术

(第2卷) (第2版)

Routing TCP/IP

Volume II
Second Edition

[美] Jeff Doyle, CCIE # 1919 著

夏俊杰 译

YESLAB工作室 审校

 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS

ciscopress.com

TCP/IP路由技术

(第2卷) (第2版)

Routing TCP/IP

Volume II
Second Edition



[美] **Jeff Doyle**, CCIE # 1919 著

夏俊杰 译

YESLAB工作室 审校

人民邮电出版社

北京

图书在版编目 (C I P) 数据

TCP/IP路由技术 : 第2版. 第2卷 / (美) 杰夫·多伊尔 (Jeff Doyle) 著 ; 夏俊杰译. -- 北京 : 人民邮电出版社, 2017.8
ISBN 978-7-115-46100-1

I. ①T… II. ①杰… ②夏… III. ①计算机网络—通信协议—路由选择 IV. ①TN915.05

中国版本图书馆CIP数据核字(2017)第156656号

版权声明

Routing TCP/IP, Volume II, Second Edition (ISBN: 9781587054709)

Copyright © 2017 Pearson Education, Inc.

Authorized translation from the English language edition published by Cisco Press.

All rights reserved.

本书中文简体字版由美国 Pearson Education 授权人民邮电出版社出版。未经出版者书面许可, 对本书任何部分不得以任何方式复制或抄袭。

版权所有, 侵权必究。

-
- ◆ 著 [美] Jeff Doyle
译 夏俊杰
审 校 YESLAB 工作室
责任编辑 傅道坤
责任印制 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>
北京鑫正大印刷有限公司印刷
- ◆ 开本: 787×1092 1/16
印张: 48.25
字数: 1211 千字 2017 年 8 月第 1 版
印数: 1-2 400 册 2017 年 8 月北京第 1 次印刷
- 著作权合同登记号 图字: 01-2016-2071 号

定价: 148.00 元

读者服务热线: (010) 81055410 印装质量热线: (010) 81055316

反盗版热线: (010) 81055315

广告经营许可证: 京东工商广登字 20170147 号

内容提要

本书是有关 Cisco 外部路由协议和高级 IP 路由主题的权威指南，是 Cisco 路由与交换领域实属罕见的经典著作。本书在上一版的基础上进行了全面更新，其可读性、广度和深度相较于上一版有了相当大的改进。

本书主要分为 11 章，其内容包括域间路由概念、BGP 简介、BGP 和 NLRI、BGP 和路由策略、扩展 BGP、多协议 BGP、IP 组播路由简介、协议无关组播、扩展 IP 组播路由、IPv4 到 IPv4 的网络地址转换（NAT44）、IPv6 到 IPv4 的网络地址转换（NAT64）等。为了方便读者深入掌握各章所学知识，本书提供了大量的案例分析材料，涵盖了协议配置、故障检测和排除等方面。每章在结束时都提供大量的复习题、配置练习和排错练习，以加强读者对所学知识的理解与记忆。

本书除了适合众多备考的准 CCIE 以及需要通过再认证的 CCIE 阅读，还非常适合从事大型 IP 网络规划、设计和实施工作的工程技术人员及网络管理员参考。

关于作者

Jeff Doyle (CCIE #1919) 是 Fishtech 实验室的研发副总裁，主要研究方向是 IP 路由协议、SDN/NFV、数据中心架构、MPLS 以及 IPv6 技术。Jeff 设计或协助设计完成的大规模 IP 服务提供商网络以及企业网络遍及六大洲的 26 个国家，曾经协助日本、中国以及韩国开展 IPv6 的早期部署，为这些国家的服务提供商、政府机构、军队供应商、设备制造商以及大型企业提供 IPv6 最佳部署方案的咨询服务。他目前主要为大型企业提供数据中心基础设施、SDN 以及 SD-WAN 等领域的演进咨询服务。

Jeff 是《TCP/IP 路由技术》(第 1 卷和第 2 卷)以及《OSPF 和 IS-IS 详解》的作者，是 *Software Defined Networking: Anatomy of OpenFlow* 一书的合著者，同时还是 *Juniper Networks Routers: The Complete Reference* 的编辑及特约作者。此外，Jeff 还为福布斯、*Network World* 博客及 *Network Computing* 博客写文章。Jeff 是洛基山 IPv6 任务组的奠基人之一（是 IPv6 论坛会员），并在 ISOC（互联网协会）科罗拉多分会的执行委员会任职。

Jeff 和他的妻子 Sara 以及一只名叫 Max 的牧羊犬住在科罗拉多州的威斯敏斯特。Jeff 和 Sara 的生活非常美满，长期与四个成年子女及一大群孙子孙女们居住在方圆几公里的范围之内。

关于特约作者

Khaled W. Abuelenain (CCIE #27401) 目前是 Acuitive 公司 (Cisco 认证的专家级可管理服务合作伙伴) 的咨询总监, 工作地点位于公司在沙特阿拉伯的 EMEA 办事处。Khaled 获得了双 CCIE (R&S 和 SP) 认证, 拥有埃及艾因·夏姆斯大学电子与通信工程专业的学士学位, 从 1997 年以来一直是 IEEE 会员。Khaled 在中东拥有 14 年以上的大型网络设计、运维及优化经验, 特别是为跨国服务提供商和移动运营商以及银行、政府机构提供服务。Khaled 在路由、BGP、MPLS 和 IPv6 等领域造诣精深, 同时还是数据中心技术以及网络可编程领域的专家, 对于 SDN 解决方案中的 Python 编程技术尤为感兴趣。他还是云计算以及 SDN IEEE 协会的活跃成员。

Nicolas Michel (CCIE #29410, R&S 和 DC 双 CCIE) 是一名网络架构师, 在路由与交换、数据中心以及统一通信等领域拥有 10 多年的工作经验。Nicolas 曾经是法国空军的前空军中士, 服役期间担任的工作角色是网络工程师, 参与了多个与北约相关的项目。

Nicolas 自 2011 年开始搬到瑞士, 为当地一家领先的网络咨询公司工作。

Nicolas 是 UEFA EURO 2016 年欧洲足球锦标赛的首席 UC 架构师。

Nicolas 喜欢研究各类网络新技术(如 SDN、自动化/网络可编程), 他的博客是 <http://vpackets.net>。Nicolas 还是一家旨在帮助自闭症儿童的非政府组织的负责人。

Nicolas 参加了一个开源的网络仿真项目: <http://www.unetlab.com/>。

目前 Nicolas 正打算移民到美国。

谨将本书献给我挚爱的妻子, 感谢她一直以来对我职业生涯的无私支持, 让我在工程师道路上不断前行, 没有她就没有今天的我。

同时还要将本书献给我的孩子和我的父母, 他们教会我永不放弃, 并且享受生活的每一刻。

最后, 衷心感谢 Jeff Doyle, 让我有机会参与本书的写作过程, 我从中学到了很多东西, 直到现在也不敢相信我是如此的幸运。

——Nicolas

关于技术审稿人

Darien Hirotsu 是一名经验丰富的网络专家，在服务提供商、数据中心以及企业网等领域拥有十多年的工作经验。**Darien** 拥有加州大学圣克鲁斯分校网络工程的硕士学位以及加州州立理工大学圣路易斯奥比斯珀分校的学士学位，同时还获得了多项专家级认证，对 SDN 中的软件及网络技术非常感兴趣。

Darien 希望在此对他的未婚妻 **Rebecca Nguyen** 表示衷心的感谢。虽然我在编辑本书的过程中受益良多，但也极为耗时，在此感谢 **Rebecca** 在整个编辑过程以及漫长的周末时光中，给予我始终如一的爱、支持与耐心，感谢你为我所做的一切，**Rebecca**！

Peter Moyer 是一名经验丰富的 IP/MPLS 咨询工程师，近年的研究兴趣是 SDN。**Peter** 在 IP 网络互连领域拥有多年厂商工作经验，在本世纪初获得了 JNCIE 认证，在 20 世纪 90 年代后期获得了 CCIE 认证。**Peter** 是多本 IP 及 SDN 网络书籍的合著者及技术编辑，而且还发表了大量与网络主题相关的论文及博客。他目前主要研究大规模数据中心和服务提供商网络（包括教育科研网络领域）。**Peter** 拥有马里兰大学的 CMIS 学士学位。

献辞

谨将本书献给我的妻子 Sara，以及我的孩子们 Anna、Carol、James 及 Katherine，同时还献给我的孙子孙女们 Claire、Caroline 及 Sam，他们是我的避风港，让我始终保持理智、谦逊和快乐。

致谢

技术书籍的作者就像是一群才华横溢的专业人士的名誉牵头人，本书也毫不例外，就像大家在接受奥斯卡大奖时的演说一样，我也要感谢诸多人士。

首先感谢 Khaled Abu El Enian 和 Nicolas Michel，他们为本书每章的最后增加了很多新的配置示例及排错练习题。此外，Khaled 还帮我节约了大量写作时间，编写了第 5 章“扩展 BGP 功能”小节的大部分内容。衷心希望在未来的写作中能够与他们保持更加密切的合作关系。

还要感谢我的好友兼同事 Pete Moyer，他不但是我独自编写的每本书的技术审稿人，而且还与我一同编写了多本图书，Peter 对我生命的影响已远远超出本书以及其他图书，我将永远心怀感激。

Darien Hirotsu 是本书的另一名技术审稿人，虽然我们在很多企业及工程项目上有过大量合作，但这一次是我们在写作上的首次合作。Darien 对于细节有着异乎常人的把握能力，帮我找出了手稿中的大量细节错误。

感谢 Chris Cleveland，感谢他作为开发编辑所提供的大量专业指导。我们一起合作了多本图书，是他让我的每一本图书都更加精益求精，也让我成为了一名更加优秀的作者。

感谢 Brett Bartow 以及 Cisco Press 的全体同仁，感谢 Brett 在本书写作期间表现出来的极大耐心，Brett 是我写作进度经常滞后的受害者，每天都得将进度控制作为日常工作的头等大事。在我认为他应该拿着本书第 1 卷敲打我脑袋的时候，他仍然一如既往地表现出了莫大的友善之情。

感谢我的妻子 Sara，多年来一直默默地陪我度过了多本图书的编写生涯。每次她看到我茫然地盯着远处时，就会说“你又开始写作啦？”

最后，还要感谢你们——我的忠实读者，是你们让这两卷 TCP/IP 路由技术变得如此成功，并一直耐心地等待我完成这个新版本，希望本书内容值得你们的期待。

前言

自从出版了《TCP/IP 路由技术（第 1 卷）》之后，虽然 Cisco Press 的“CCIE 职业发展系列”增加了大量内容，而且 CCIE 项目本身也扩展到了多个专业领域，但 IP 路由协议仍然是所有准 CCIE 们的核心基础，他们必须透彻理解和掌握，否则基础不牢，大厦将倾。

我在《TCP/IP 路由技术（第 1 卷）》的前言中曾经说过“……随着互连网络规模和复杂性的不断增大，路由问题也将立刻变得庞大且错综复杂”，由于本书重点从 IGP 转移到了自治系统间的路由问题以及组播和 IPv6 等诸多特殊路由问题，因而可扩展性和管理性仍然是本书第 2 卷的核心主题。

本书的写作目的不仅是要帮助读者轻松通过 CCIE 实验室考试，从而在名字后面加上极具价值的 CCIE 编号，而且希望帮助读者不断增进知识与技巧，从而无愧于 CCIE 称号。正如我在《TCP/IP 路由技术（第 1 卷）》中曾经说过的一样，我希望读者成为真正的 CCIE，而不仅仅是一名能够通过 CCIE 实验室考试的人员，因而本书所提供的信息要远远多于通过实验室考试所需的知识。当然，所有信息对于一名受人尊敬的互连网络专家的职业生涯来说都是至关重要的。

在我获得 CCIE 称号时，CCIE 实验室还主要是由 AGS+路由器组成的，与那个时代相比，现在的 CCIE 实验室和考试内容已经发生了翻天覆地的变化，当前实验室考试的难度变得越来越高，而且 CCIE 项目还增加了再认证要求。在我第一次参加再认证考试之前，有人曾告诉过我《TCP/IP 路由技术（第 1 卷）》对他们准备该考试起到了巨大的作用，特别是 IS-IS（该协议几乎没有用在服务提供商的网络之外）。因而我决定写作本书的第 2 卷，不仅面向众多准 CCIE 们，而且也面向那些需要通过再认证的 CCIE 们。有关组播路由及 IPv6 的章节就是面向这样的读者群的。

我努力遵照《TCP/IP 路由技术（第 1 卷）》的结构来编写第 2 卷，即首先从通用角度描述某种协议，然后以 Cisco IOS 为例给出相应的协议配置示例，最后再给出利用 Cisco IOS 工具检测与排除协议故障的示例，但对于 BGP 和 IP 组播来说，如果按照这种结构来编写，那么将会使单章内容变得极为冗长，因而我将其分解成了多个章节。

最后，衷心希望大家阅读本书时获得的知识丝毫不亚于我写作本书所获得的知识。

第 2 版前言

几乎在本卷第 1 版于 2001 年首次发行之，我就希望增加和修改某些内容，主要原因来自于我不断积累的工作经验。从 1998 年到 2010 年，我的工作对象基本上都是服务提供商和运营商，从这些设计项目、技术决策以及主导或参与的众多技术交流中，我学到了很多新知识，当然，有些新知识仅仅弥补了我个人经验上的不足，但并不是所有的新知识都是如此。在 BGP 及组播网络变得越来越复杂、涌现出大量新功能且最佳实践也在不断发展变化的情况下，我也一直在与业界的发展变化保持同步。

业界发生了哪些变化

下面将简要描述本书第一版发行之，业界发生的一些新变化。

BGP

有关 BGP 的主要概念都已经在 2001 年发行的本书第 1 版中做了详细描述, BGP 是一种被互联网广泛使用的外部网关协议(即自治系统间路由协议), 具备多协议处理能力, 版本 4 是目前可接受的版本。虽然这些年 BGP 也增加了一些新的功能特性和协议能力, 但协议本身并没有出现大的变化。

主要变化之处在于业界对 BGP 的使用经验, 这些经验不但增强了人们使用 BGP 策略的方式, 而且在某些情况下还改变了传统的最佳实践。此外, 多协议 BGP 已成为多业务核心网的主力, 由于多协议 BGP 允许定义大量新的地址簇, 因而可以在单个共享的核心网上运行多种不同的业务。虽然本书并没有讨论多业务网的另一个必要因素——MPLS (Multiprotocol Label Switching, 多协议标签交换), 这是因为有关 MPLS 的内容完全可以单出一本或两本专著, 但读者完全可以通过本书介绍的这些多协议 BGP 知识, 理解多协议 BGP 对于各种基于 MPLS 的地址簇的支持方式。此外, 本书还提供了大量配置示例, 以帮助读者正确理解多协议 BGP 在 IPv4 和 IPv6 下支持单播和组播地址簇的方式。

本书第 1 版安排了一个章节专门介绍 EGP (BGP 的前身), 虽然那时已经废止了 EGP, 但某些政府网络仍在使用该协议, 这也是本书在第一版仍然涵盖 EGP 的主要原因之一, 另一个主要原因就是防止某些不循常理的实验室考官在 CCIE 考试中突然抛出一些 EGP 考题。考虑到目前该协议已基本绝迹, 因而第 2 版仅在介绍 BGP 时将 EGP 作为背景知识进行简要交代。

为了反映业界在 BGP 使用经验上的不断丰富以及 Cisco 新支持的大量新 BGP 功能特性, 本书第 2 版将第 1 版中有关 BGP 的 2 章内容扩展到了 6 章。

IP 组播

IP 组播网络的发展变化可能比 BGP 网络的发展变化更大, 由于组播及其相关联的路由协议极其复杂, 因而在 2001 年的时候还很难管理组播网络。虽然从某种意义上来说, 这些困难目前依然存在, 但业界出现的一些变化使得这些困难不再高不可攀。

虽然 2001 年最常见的组播路由协议是 DVMRP、PIM-DM 和 PIM-SM, 但当时我推断 CBT (Core-Based Tree, 核心树) 和 MOSPF (Multiprotocol OSPF, 多协议 OSPF) 可能会成为主流, 因而在第 1 版中介绍了这方面的内容, 不过从目前来看, CBT 和 MOSPF 一直未被接受, DVMRP 也成了组播路由协议中的 RIP (已被废止, 但在某些场合依然能够看到), 因而在第 2 版中删除了有关 CBT 和 MOSPF 的全部内容, 仅做简单交代, 而且与第 1 版相比, 有关 DVMRP 的介绍也做了大幅简化。

由于 PIM 已成为当下 IPv4 和 IPv6 网络广泛接受的组播路由协议, 因而本书第 2 版更加深入地介绍了有关 PIM-DM、PIM-SM 以及 PIM-SSM 的内容。

IPv6

虽然我从 1990 年代后期就一直倡导和推广 IPv6, 但截至 2001 年的时候, 对这个新版本 IP 协议感兴趣的国家和地区仅限于日本、中国和韩国, 美国和欧盟则毫不关心(少数军事领域除外), 他们认为 IPv6 在很大程度上只是面向未知的将来, 那时候所有预测公有 IPv4 地址池将在 2012 年耗尽的人都被认为是杞人忧天, 显得荒谬可笑。因而我在本书第 1 版单独安排了一章讨论 IPv6, 与书中其他主题几乎毫无关系。

但这 15 年确实是天翻地覆的 15 年，目前 IPv6 已成为当前的主流协议，估计要不了几年 IPv6 就将全面替代目前已经耗尽的 IPv4。为了反映当前的实际情况，第 2 版不再将 IPv6 单独列为一个章节，而是将 IPv6 的支持要求贯穿于整个 BGP 及 IP 组播的讨论当中。

2001 年的网络地址转换指的是 NAT-PT，一般仅在不同的 IPv4 地址之间进行转换，十几年来网络地址转换技术得到了极大扩展，因而第 2 版安排了两章内容来讨论 NAT：一章讨论 IPv4 到 IPv4 的地址转化；另一章则讨论 IPv6 到 IPv4 的地址转换（NAT64）。

第 2 版有哪些变化

第 2 版在章节安排上的最后一个差异就是去掉了第 1 版中关于路由器管理的章节（第 1 版中的第 9 章），这是因为 2001 年之后有关 Cisco 路由器管理的主题变得越来越庞大，Cisco 也提供了越来越多的路由平台，必须花费大量篇幅才有可能解释清楚，但这与本书的主旨相悖，而且本书的名字毕竟是 TCP/IP 路由技术，而不是 Cisco 路由器管理技术。

第 2 版的其他变化如下。

- **IOS:** IOS 是 2001 年唯一的 Cisco 路由器操作系统，目前除了 IOS 之外，还有 IOS-XR、IOS-XE 和 NX-OS，要想完全覆盖这些操作系统的配置示例及配置练习，不但极为繁琐和复杂，而且还与两卷 TCP/IP 路由技术的主要目标（讲解协议相关内容）相悖，因而本书仅以 IOS 为例。理解了 IOS 之后，读者完全可以很轻松地理解其他 Cisco 操作系统。
- **Cisco 与 IOS:** 与前一项有关，第 1 版通常使用“Cisco 命令”的表述方式，考虑到目前 Cisco 提供了多种操作系统，因而第 2 版尽量准确地使用“IOS 命令”的表述方式。
- **命令与语句:** 仍然与前一项有关，第 2 版尽量区分 IOS 命令与 IOS 语句。第 2 版中的命令表示输入某些信息之后期望得到直接结果，而语句则属于 IOS 配置的一部分，影响路由器的运行状态。
- **命令参考:** 第 1 版在每章的最后都以列表方式列出了本章使用过的所有命令（和语句），给出这些命令的完整语法格式与描述信息。由于第 2 版包含的命令（和语句）过多，而且不同 IOS 版本有时还存在不同的语法格式，导致这张表格过于冗长，也不实用，因而第 2 版的每章最后不再提供命令参考，如果希望了解某个命令（或语句）的完整语法信息，可以在线查询相应 IOS 版本的“Cisco Command Reference Guide”，查询时请要注意 IOS 版本。
- **IOS 版本:** 第 1 版的大部分示例均使用 IOS 11.4，如前所述，目前的 IOS 版本非常多，虽然某些示例仍然使用了第 1 版的部分输出结果，但大多数情况下使用的都是最新的 12.4 或 15.0。对于所有示例来说，本书保证所提供的配置信息或输出结果完全包含所讨论的信息，不过读者实际看到的输出结果可能与书中示例并不完全一致，这取决于读者使用的 IOS 版本，因而请读者重点关注输出结果中的有用信息，而不要过多关心输出结果是否与实际看到的输出结果完全一致。
- **集成式故障检测与排除:** 《TCP/IP 路由技术（第 1 卷）》中的每一章最后都安排了一些固定内容，包括对每章主题的简要技术概述、IOS 配置练习题以及故障检测与排除练习题。考虑到第 2 版中的 BGP 和 IP 组播技术都非常复杂，因而在组织这两部分内容时，将故障检测与排除案例都集成到通用配置示例中了。

- **网络与互连网络**：这是一个非常细微的变化。2001 年的时候我曾试图准确定义这两个概念，网络指的是一种常规通信介质（如以太网），而互连网络指的是由路由器互连起来的多个网络。从目前来看，这是一种过时的说法，因为目前几乎无人再提互连网络（少数严谨场合除外），因而第二版删除了所有的互连网络一词。与共享通信介质相比，子网的逻辑含义以及与地址相关的含义更加丰富，网络一词需要从使用该词的上下文来加以理解，因而网络可能表示由一条串行链路或以太网链路互连的两台路由器，也可能表示一个巨大的 AS 间系统（如互联网）。虽然不是很严谨，但路由器工程师们每天都在茶余饭后使用这个词。
- **怪异的之字形串行链路图标**：从我在 20 世纪 90 年代早期讲授 Cisco 课程开始，就一直使用之字形线或“闪电形”线来表示串行链路，如此区分的原因在于串行链路的特性与 LAN 链路存在差异，但是对于本卷图书的所有示例来说，链路类型与示例并无任何关系，而且我发现之字形图标经常会让插图显得凌乱不堪，因而我尽量将书中用于接口间互连的所有之字形图标全部更换为直线，而不管这些链路的类型是什么。

配置练习题和故障检测与排除练习题答案

读者需要下载两个附录以查看配置练习题和故障检测与排除练习题的答案：附录 B 和附录 C。

读者可在 www.epubit.com.cn/book/details/4061 的页面上下载这两个附录。

命令语法定义

本书在介绍命令语法时使用与 IOS 命令参考一致的约定，本书涉及的命令参考约定如下：

- 需要逐字输入的命令和关键字用**粗体**表示，在配置示例和输出结果（而不是命令语法）中，需要用户手工输入的命令用**粗体**表示（如 **show** 命令）；
- 必须提供实际值的参数用斜体表示；
- 互斥元素用竖线（|）隔开；
- 中括号[]表示可选项；
- 大括号表示{}必选项；中括号内的大括号[{}]表示可选项中的必选项。

目 录

第 1 章 域间路由概念	1
1.1 早期域间路由协议: EGP	1
1.1.1 EGP 起源	1
1.1.2 EGP 操作	3
1.1.3 EGP 的不足	11
1.2 BGP 的出现	12
1.3 BGP 基础	12
1.4 自治系统类型	15
1.5 EBGP 与 IBGP	16
1.6 多归属	22
1.6.1 转接 AS 多归属	22
1.6.2 末梢 AS 多归属	23
1.6.3 多归属与路由策略	27
1.6.4 多归属面临的问题: 负载共享与负载均衡	27
1.6.5 多归属面临的问题: 流量控制	28
1.6.6 多归属面临的问题: PA 地址	30
1.7 CIDR	31
1.7.1 汇总概述	31
1.7.2 无类别路由	32
1.7.3 汇总: 好处、坏处及 不对称流量	36
1.7.4 CIDR: 延缓 B 类地址 空间的耗尽速度	38
1.7.5 CIDR: 降低路由表爆 炸的风险	38
1.7.6 管理和分配 IPv4 地址块	41
1.7.7 CIDR 面临的问题: 多归属与 PA 地址	43
1.7.8 CIDR 面临的问题: 地址可携带性	44
1.7.9 CIDR 面临的问题: PI 地址	45
1.7.10 CIDR 面临的问题: 流量工程	45
1.7.11 CIDR 的问题解决之道	47
1.7.12 IPv6 时代的到来	50
1.7.13 再论 Internet 路由表爆炸	50
1.8 展望	52
1.9 复习题	52
第 2 章 BGP 简介	53
2.1 谁需要 BGP	53
2.1.1 连接非信任域	53
2.1.2 连接多个外部邻居	54
2.1.3 设置路由策略	58
2.1.4 BGP 的危害	61
2.2 BGP 操作	62
2.2.1 BGP 消息类型	63
2.2.2 BGP 有限状态机	64
2.2.3 路径属性	67
2.2.4 BGP 决策进程	74
2.2.5 BGP 消息格式	76
2.2.6 Open 消息	77
2.2.7 Update 消息	78
2.2.8 Keepalive 消息	80
2.2.9 Notification 消息	80
2.3 BGP 对等关系的配置及故障 检测与排除	81
2.3.1 案例研究: EBGP 对等会话	81
2.3.2 案例研究: 基于 IPv6 的 EBGP 对等会话	84
2.3.3 案例研究: IBGP 对等会话	87
2.3.4 案例研究: 直连检查与 EBGP 多跳	93
2.3.5 案例研究: 管理和保护 BGP 连接	99

2.4	展望	103	4.2.1	InQ 与 OutQ	175
2.5	复习题	104	4.2.2	IOS BGP 进程	177
2.6	配置练习题	104	4.2.3	NHT、Event 以及 Open 进程	180
2.7	故障检测与排除练习题	105	4.2.4	表版本	182
第 3 章	BGP 与 NLRI	111	4.3	管理策略变更	188
3.1	在 BGP 中配置 NLRI 以及 检测与排除 NLRI 故障	111	4.3.1	清除 BGP 会话	188
3.1.1	利用 network 语句 注入前缀	111	4.3.2	软重配	189
3.1.2	利用 network mask 语句注入前缀	114	4.3.3	路由刷新	192
3.1.3	利用重分发注入前缀	115	4.4	路由过滤技术	196
3.2	NLRI 与 IBGP	119	4.4.1	通过 NLRI 过滤路由	196
3.2.1	在 IBGP 拓扑结构中 管理前缀	120	4.4.2	案例研究: 使用分发列表	197
3.2.2	IBGP 与 IGP 同步	128	4.4.3	使用扩展 ACL 的路由 过滤器	205
3.3	将 BGP NLRI 宣告到 本地 AS 中	129	4.4.4	案例研究: 使用前缀列表	206
3.3.1	将 BGP NLRI 重分 发到 IGP 中	130	4.4.5	使用 AS_PATH 过滤路由	213
3.3.2	案例研究: 利用 IBGP 将 NLRI 重分发到末梢 AS 中	130	4.4.6	正则表达式	213
3.3.3	利用静态路由将 NLRI 宣 告到末梢 AS 中	137	4.4.7	案例研究: 使用 AS_PATH 过滤器	217
3.3.4	将默认路由宣告给 邻接 AS	139	4.4.8	案例研究: 利用路由映射 设置策略	220
3.4	利用 BGP 宣告聚合路由	140	4.4.9	过滤器处理	225
3.4.1	案例研究: 利用静态 路由进行聚合	141	4.5	影响 BGP 决策进程	226
3.4.2	利用 aggregate-address 语句进行聚合	142	4.5.1	案例研究: 管理权重	227
3.4.3	ATOMIC_AGGREGATE 与 AGGREGATOR 属性	146	4.5.2	案例研究: 使用 LOCAL_ PREF 属性	234
3.4.4	聚合时使用 AS_SET	149	4.5.3	案例研究: 使用 MULTI_ EXIT_DISC 属性	240
3.5	展望	154	4.5.4	案例研究: 附加 AS_PATH	256
3.6	复习题	155	4.5.5	案例研究: 管理距离与 后门路由	260
3.7	配置练习题	155	4.6	控制复杂的路由映射	265
3.8	故障检测与排除练习题	158	4.6.1	continue 子句	266
第 4 章	BGP 与路由策略	167	4.6.2	策略列表	268
4.1	策略与 BGP 数据库	168	4.7	展望	270
4.2	IOS BGP 实现	175	4.8	复习题	270
			4.9	配置练习题	271
			4.10	故障检测及排除练习题	274
			第 5 章	扩展 BGP	281
			5.1	扩展配置	282
			5.1.1	对等体组	282
			5.1.2	对等体模板	289
			5.1.3	COMMUNITY 属性	297

5.2 扩展 BGP 功能.....	334	第 7 章 IP 多播路由简介.....	496
5.2.1 路由翻动抑制.....	334	7.1 IP 多播需求.....	498
5.2.2 ORF.....	339	7.1.1 IPv4 多播地址.....	499
5.2.3 NHT.....	346	7.1.2 IPv6 多播地址.....	502
5.2.4 快速外部切换.....	355	7.1.3 组成员概念.....	504
5.2.5 BFD.....	357	7.1.4 IGMP.....	508
5.2.6 BGP PIC.....	365	7.1.5 MLD.....	517
5.2.7 GR.....	376	7.1.6 IGMP/MLD Snooping.....	520
5.2.8 最大前缀.....	377	7.1.7 CGMP.....	522
5.2.9 调节 BGP CPU.....	386	7.2 多播路由的问题.....	525
5.2.10 调节 BGP 内存.....	388	7.2.1 多播转发.....	526
5.2.11 BGP 传输优化.....	393	7.2.2 多播路由.....	527
5.3 扩展 BGP 网络.....	398	7.2.3 稀疏与密集拓扑结构.....	528
5.3.1 私有 AS 号.....	398	7.2.4 隐式加入与显式加入.....	529
5.3.2 4 字节 AS 号.....	402	7.2.5 有源树与共享树.....	530
5.3.3 IBGP 与 N 平方问题.....	402	7.2.6 SSM.....	531
5.3.4 联盟.....	403	7.2.7 多播定界.....	532
5.3.5 路由反射器.....	414	7.3 展望.....	535
5.4 展望.....	424	7.4 推荐读物.....	535
5.5 复习题.....	425	7.5 复习题.....	535
5.6 配置练习题.....	426	7.6 配置练习题.....	536
5.7 故障检测及排除练习题.....	428	第 8 章 PIM.....	538
第 6 章 多协议 BGP.....	430	8.1 PIM 简介.....	538
6.1 BGP 的多协议扩展.....	430	8.2 PIM-DM 操作.....	539
6.2 MBGP 支持 IPv6 地址簇.....	432	8.2.1 PIM-DM 基础.....	539
6.3 配置 IPv6 MBGP.....	433	8.2.2 剪除覆盖.....	544
6.3.1 IPv4 TCP 会话上的 IPv4 和 IPv6 前缀.....	434	8.2.3 单播路由变化.....	545
6.3.2 将 IPv4 BGP 配置更 新为地址簇格式.....	440	8.2.4 PIM-DM 指派路由器.....	545
6.3.3 IPv6 TCP 会话上的 IPv4 和 IPv6.....	442	8.2.5 PIM 转发路由器选举.....	546
6.3.4 双栈 MBGP 连接.....	449	8.3 PIM-SM 操作.....	548
6.3.5 多跳双栈 MBGP 连接.....	453	8.3.1 PIM-SM 基础.....	549
6.3.6 IPv4 和 IPv6 混合会话.....	454	8.3.2 发现 RP.....	549
6.3.7 多协议 IBGP.....	457	8.3.3 PIM-SM 与共享树.....	554
6.3.8 案例研究：多协议 策略配置.....	465	8.3.4 源注册.....	556
6.4 展望.....	491	8.3.5 PIM-SM 与最短路径树.....	561
6.5 复习题.....	491	8.3.6 PIMv2 消息格式.....	565
6.6 配置练习题.....	492	8.4 IP 多播路由的配置.....	572
6.7 故障检测及排除练习题.....	494	8.4.1 案例研究：配置 PIM-DM.....	573
		8.4.2 案例研究：配置 PIM-SM.....	579
		8.4.3 案例研究：多播负载均衡.....	598
		8.5 IP 多播路由的故障检测 与排除.....	603

8.5.1 使用 mrinto	604	10.2.4 安全	661
8.5.2 使用 mtrace 和 mstat	606	10.2.5 协议相关问题	661
8.6 展望	609	10.3 配置 NAT44	668
8.7 推荐读物	609	10.3.1 案例研究: 静态 NAT	668
8.8 复习题	610	10.3.2 NAT44 与 DNS	673
8.9 配置练习题	610	10.3.3 案例研究: 动态 NAT	674
8.10 故障检测与排除练习题	612	10.3.4 案例研究: 网络融合	678
第 9 章 扩展 IP 多播路由	615	10.3.5 案例研究: 通过 NAT 多归属到 ISP	682
9.1 多播定界	615	10.3.6 案例研究: PAT	686
9.2 案例研究: 多播穿越 非多播域	618	10.3.7 案例研究: TCP 负载 均衡	687
9.3 连接 DVMP 网络	620	10.3.8 案例研究: 服务分发	689
9.4 AS 间多播	622	10.4 NAT44 的故障检测与排除	690
9.4.1 MBGP	624	10.5 展望	692
9.4.2 MSDP 操作	625	10.6 复习题	692
9.4.3 MSDP 消息格式	627	10.7 配置练习题	692
9.5 案例研究: 配置 MBGP	630	10.8 故障检测与排除练习题	694
9.6 案例研究: 配置 MSDP	634	第 11 章 NAT64	696
9.7 案例研究: MDSP 网状 多播组	638	11.1 SIIT	697
9.8 案例研究: 任播 RP	640	11.1.1 IPv4/IPv6 报头转换	699
9.9 案例研究: MSDP 默认 对等体	644	11.1.2 ICMP/ICMPv6 转换	700
9.10 展望	646	11.1.3 分段与 PMTU	703
9.11 复习题	646	11.1.4 上层报头转换	704
9.12 配置练习题	647	11.2 NAT-PT	704
第 10 章 NAT44	650	11.2.1 NAT-PT 操作	705
10.1 NAT44 操作	650	11.2.2 配置 NAT-PT	707
10.1.1 NAT 的基本概念	651	11.2.3 为什么要废止 NAT-PT	720
10.1.2 NAT 与节约 IP 地址	652	11.3 无状态 NAT64	721
10.1.3 NAT 与 ISP 迁移	654	11.3.1 无状态 NAT64 操作	722
10.1.4 NAT 与多归属自治系统	655	11.3.2 配置无状态 NAT64	724
10.1.5 PAT	657	11.3.3 无状态 NAT64 的局限性	726
10.1.6 NAT 与 TCP 负载分发	658	11.4 有状态 NAT64	726
10.1.7 NAT 与虚拟服务器	659	11.4.1 有状态 NAT64 操作	726
10.2 NAT 问题	660	11.4.2 配置有状态 NAT64	728
10.2.1 报头检验	660	11.4.3 有状态 NAT64 的局限性	730
10.2.2 分段	660	11.5 展望	730
10.2.3 加密	661	11.6 复习题	730
		11.7 配置练习题	731
		附录 A 复习题答案	733

第 1 章

域间路由概念

如果所有网络都使用单一路由协议（如 OSPF 或 IS-IS），那么可以想象如今的互联网会是什么样子。如果每个子网地址均可见的情况下，那么将根本无法保证网络的稳定性。网络的安全性也将脆弱不堪，这是因为针对路由协议的攻击（甚至是一个不起眼的配置差错）都可能会导致整个互联网的瘫痪。此外，谁能管理这样的网络呢？如何在全世界范围内协调所有网络管理员开展协议升级或协议增强等繁琐的维护工作呢？

随着 ARPANET（现代互联网的前身）的规模在 20 世纪 70 年代后期变得越来越庞大，上述问题也就接踵而至，人们开始尝试创建一种可扩展的方式来管理网络。当时最基本的思路就是创建称为 AS（Autonomous System，自治系统）的管理域（AS 定义了同一管理权限下的网络边界）以及可在这些管理域之间进行路由的协议（该路由协议对于管理域来说是外部路由协议）。

目前用于自治系统间的路由协议是 BGP（Border Gateway Protocol，边界网关协议）。本书将在第 2 章到第 6 章描述 BGP 的功能、配置、故障检测与排除以及各种相关的路由策略。在讨论 BGP 之前，本书将在第 1 章介绍 BGP 背后的关键概念及其演进过程。换句话说，第 2 章到第 6 章解释的是如何工作的问题，而第 1 章解释的则是能做什么以及为什么的问题。

1.1 早期域间路由协议：EGP

本书第一版花了整章篇幅来解释 BGP 的前身——EGP（Exterior Gateway Protocol，外部网关协议）。那时的 EGP 基本处于已经废止状态，只有少数老旧网络仍在使用 EGP。几年时间之后，EGP 已经被完全废止，IOS 也不再支持该协议。

不过 EGP 可以帮助读者理解当初在设计可操作的域间（即 Inter-AS）路由协议时的思路。本节将从技术历史的角度简述 EGP 的发展历程，以便更好地理解 BGP 的发展过程。虽然读者也可以选择略过本节，但本节介绍的某些基本概念将一直沿用到 BGP 上，而且 EGP 的某些功能还有助于理解 BGP 设计思路背后的一些概念。

1.1.1 EGP 起源

在 20 世纪 80 年代早期，组成 ARPANET 的路由器（网关）设备都运行了一种距离向量