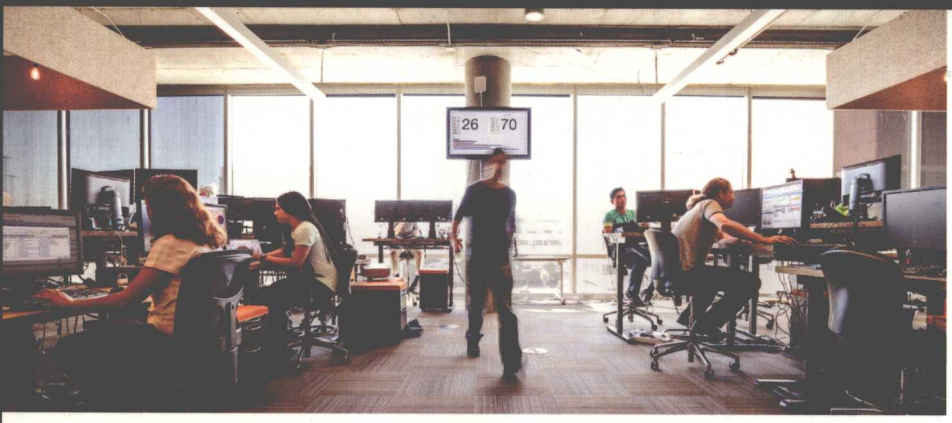
CISCO

ciscopress.com



IP组播（第2卷）

IP Multicast, Volume II

Advanced Multicast Concepts and
Large-Scale Multicast Design

[美] 乔西·洛夫莱斯（Josh Loveless）
雷·布莱尔（Ray Blair） 著
阿温德·杜莱（Arvind Durai）

古宏霞 王涛 宣建国 孙余强 译

中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS

ciscopress.com

IP组播 (第2卷)

IP Multicast, Volume II

Advanced Multicast Concepts and Large-Scale Multicast Design

[美] 乔西·洛夫莱斯（Josh Loveless）
雷·布莱尔（Ray Blair） 著
阿温德·杜莱（Arvind Durai）

古宏霞 王涛 宣建国 孙余强 译

人民邮电出版社

北京

图书在版编目（CIP）数据

IP组播. 第2卷 / (美) 乔西·洛夫莱斯
(Josh Loveless)，(美) 雷·布莱尔 (Ray Blair)，
(美) 阿温德·杜莱 (Arvind Durai) 著；古宏霞等译
. -- 北京：人民邮电出版社，2018.8
ISBN 978-7-115-48675-2

Ⅰ. ①I… Ⅱ. ①乔… ②雷… ③阿… ④古… Ⅲ. ①
互连网络－通信协议 Ⅳ. ①TN915.04

中国版本图书馆CIP数据核字(2018)第129283号

版权声明

◆ 著　[美] 乔西·洛夫莱斯（Josh Loveless）
[美] 雷·布莱尔（Ray Blair）
[美] 阿温德·杜莱（Arvind Durai）
译　古宏霞 王涛 宣建国 孙余强
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
固安县铭成印刷有限公司印刷

◆ 开本：800×1000　1/16
印张：22.5
字数：494 千字　2018 年 8 月第 1 版
印数：1－2 000 册　2018 年 8 月河北第 1 次印刷

著作权合同登记号　图字：01-2017-8639 号

定价：89.00 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

本书在第 1 卷的基础之上介绍了组播的一些高级概念外加在设计大型组播网络时用到的一些方法和技术。

本书分为 6 章，分别讲解了域间路由和 Internet 组播、组播的可扩展性以及流量传输的多样性、组播 MPLS VPN、组播之于数据中心网络环境、组播设计解决方案、排除高难度组播故障等内容。

本书适合网络工程师和架构师、网络顾问以及网络管理人员阅读，也是相关专业的读者了解 IP 组播技术的有益读物。

关于作者

Josh Loveless，CCIE #16638，Cisco 公司系统工程经理。自 2012 年起，他一直效力于 Cisco 公司，为诸多一级服务提供商（tier 1 service provider）以及 Cisco 最大的企业客户提供网络架构及支持服务，专攻大型路由和交换设计。加盟 Cisco 之前，他作为网络工程师和架构师为多家大服务提供商和大企业工作了 15 年之久，还为 Cisco 公司的某些亲密伙伴提供培训和网络架构服务。Josh 拥有路由/交换及 SP 两张 CCIE 证书。

Ray Blair，CCIE #7050，Cisco 公司杰出的系统工程师，自 1999 年以来一直效力于 Cisco 公司。他用自己多年的经验，将技术解决方案与业务需求融为一体，来确保客户业务的顺利发展。Ray 出道于 1988 年，第一份工作是设计工业监控及通信系统。从那时起，他开始参与服务器/数据库管理，以及网络的设计、实施和管理（接触了从 ATM 到 ZMODEM 在内的各种联网技术）。他拥有路由/交换、安全和 SP 三张 CCIE 证书，还是认证的信息系统安全专家（CISSP）和认证的业务架构师（证书编号 00298）。Ray 是三本 Cisco Press 图书（*Cisco Secure Firewall Services Module*、*Tcl Scripting for Cisco IOS* 以及 *IP Multicast, Volume 1*）的合著者。他在业界的许多活动中发表过演讲，还是 Cisco Live 重要的演讲者。

Arvind Durai，CCIE #7016，Cisco 高级服务团队解决方案集成部的总监。Arvind 是为 West Enterprise Region 提供高级服务的首席架构师，该组织（West Enterprise Region）大约由 100 位专家组成，致力于让 150 家左右的企业顺利发展业务。在过去的 18 年里，Arvind 一直负责支撑 Cisco 公司在企业领域里的重要客户，涉及金融行业、零售业、制造业、电子商务行业、州一级的政府机构、公共事业机构以及医疗行业等。安全、组播、网络虚拟化、数据中心企业云采用（data center enterprise cloud adoption）、自动化以及软件定义基础设施是他关注的重点。他撰写了多份白皮书，涉及各种技术。他一直为大大小小、不同行业的企业客户进行组播设计。他还是高级服务组播审计（Advanced Services Multicast Audit）工具框架的贡献者之一，客户可利用该工具来评估自己运维的组播网络，从而实现业界的最佳做法。Arvind 拥有路由/交换和安全两张 CCIE 证书，还是认证的业务架构师。他拥有电子和通信学学士学位、电子工程学硕士学位以及工商管理学硕士学位。他是 4 本 Cisco Press 图书（*Cisco Secure Firewall Services Module*、*Virtual Routing in the Cloud*、*Tcl Scripting for Cisco IOS* 和 *IP Multicast, Volume 1*）的合著者。他还与别人共同制定了 IEEE WAN 智能网格体系结构（IEEE WAN smart grid architecture），并出席过 IEEE 和 Cisco Live 等多个业界论坛。

关于技术审稿人

Nick Garner，CCIE #17871，Cisco 公司解决方案集成架构师，过去 8 年，他一直效力于 Cisco 公司高级服务团队，按交易及采购承诺为客户提供服务。他的主要职责是，为旧金山湾区的知名客户设计、部署及运维大型数据中心网络。除了数据中心路由和交换设计之外，他的主攻方向还包括安全及组播。加盟 Cisco 公司之前，Nick 在一家大型国家金融机构任网络安全工程师一职。Nick 拥有路由/交换及安全两张 CCIE 证书。

Yogeshwaran Raghunathan，CCIE #6583，Cisco 公司高级服务团队的高级解决方案集成架构师。Yogi 拥有 CIT（位于印度哥印拜陀）工商管理硕士学位以及电子和通信专业工程学学位。他在网络行业摸爬滚打了 22 年，其中有 17 年在 Cisco 公司度过，为北美的多家服务提供商提供支持。在搭建和支撑大型服务提供商网络方面所具备的实践经验，使 Yogi 认清了复杂的 MPLS 体系结构，从而对 SDN 和 MPLS 部署这一新的领域提出了不同的见解。近年来，Yogi 一直在从事大型 Web 提供商网络的设计、实施和规划工作。

献辞

谨将本书献给我的家庭以及我的朋友们，感谢你们这么多年来在工作上对我的支持。

——Josh Loveless

谨将本书献给我的妻子 Sonya 以及我的孩子 Sam、Riley、Sophie 和 Regan。你们是我的全部！

——Ray Blair

谨将本书献给我的父母和家人，感谢你们对我的支持和祝福。

——Arvind Durai

致谢

特别感谢本书的合著者 Ray Blair 和 Arvind Durai，感谢你们和我协力完成了《IP 组播》双卷本这项伟大的工作。还要感谢本书的技术审稿人 Yogi、Nick，以及 Person 出版社的所有编辑，感谢你们为了让本书大卖所做出的不懈努力！

——Josh Loveless

感谢 Josh 和 Arvind 的努力合作，Nick 和 Yogi 出彩的评论，以及 Pearson 出版社的支持。

——Ray Blair

感谢 Monica 和 Akhhill，感谢你们的耐心和一直以来对我的关照，帮我完成了我的第 5 本书。

感谢 Ray 和 Josh，感谢你们将本书第 1 卷、第 2 卷的写作变为一段欢乐之旅。

特别感谢 Brett Bartow、Yogi Raghunathan 和 Nick Garner，感谢你们提出的宝贵意见。

——Arvind Durai

前言

本书涵盖与 IP 组播设计和协议有关的高级知识，只针对 Cisco 路由器和交换机。本书包括对高级 IP 组播网络的常用特性、部署模式和现场实施的实用性讨论，并以高级 Cisco IP 组播网络的实施和故障排除所使用的命令与方法来结束讨论。

本书的读者

本书适用于任何一位负责支撑 IP 组播网络的专业人士。本书虽专为下列人士而著，但网络部门的管理者和网管人员也能从本书包括的案例研究和特性说明中受益：

- IP 网络工程师和架构师；
- 网络运维技术人员；
- 网络顾问；
- 安全专家；
- 协作专家和架构师。

本书组织结构

本书共分 6 章，涵盖以下主题。

- **第 1 章，“域间路由和 Internet 组播”**：介绍了域间组播的基本需求，外加域间组播设计的三大基本要素，即用来标识组播源主机的控制平面、用来标识组播接收主机的控制平面以及下游控制平面。
- **第 2 章，“组播的可扩展性以及流量传输的多样性”**：传输组播消息需考虑若干因素，在云服务提供商不支持客户“原生”组播的情况下，要考虑的就更多了。本章将介绍云服务的重要概念，还会解释支撑组播服务的各种要素。
- **第 3 章，“组播 MPLS VPN”**：组播 VPN 提供了一种在同一座物理基础设施之内隔离流量的功能。大多数服务提供商和诸多企业客户都部署了多协议标签交换（MPLS），以便在多个逻辑域或组（俗称虚拟专用网络[VPN]）之间分离或隔离流量。本章介绍组播 VPN 的若干实施选项。

- **第 4 章，“组播之于数据中心网络环境”**：本章将介绍数据中心网络内组播的部署方式。理解不同解决方案之间组播功能性方面的细微差异，对读者所在单位的业务顺利开展至关重要。通过介绍各种最受欢迎的数据中心实施方法（包括虚拟端口信道[Virtual Port Channel，VPC]、虚拟可扩展局域网[Virtual Extensible LAN，VXLAN]以及以应用为中心的基础设施（Application Centric Infrastructure，ACI），让读者深入领悟组播的部署方式。
- **第 5 章，“组播设计解决方案”**：本章将探讨几种典型的网络设计模型。其中一种模型展示了一种特殊的网络方案，可满足特殊的商业用途——证券交易所。另一种模型则是针对特殊行业的通用设计，侧重于医院网络环境的组播部署。本章的目的是为每一种设计类型设定一个基准，同时给出组播部署的最佳做法示例。
- **第 6 章，“排除高难度组播故障”**：本章将介绍 IP 组播网络故障排除的基本方法。

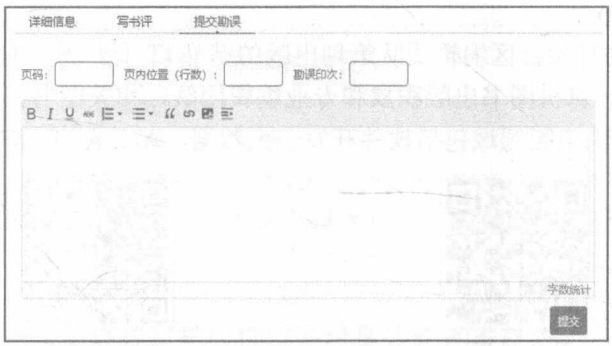
资源与支持

本书由异步社区出品，社区（https://www.epubit.com/）为您提供相关资源和后续服务。

提交勘误

作者和编辑尽最大努力来确保书中内容的准确性，但难免会存在疏漏。欢迎您将发现的问题反馈给我们，帮助我们提升图书的质量。

当您发现错误时，请登录异步社区，按书名搜索，进入本书页面，点击“提交勘误”，输入勘误信息，点击“提交”按钮即可。本书的作者和编辑会对您提交的勘误进行审核，确认并接受后，您将获赠异步社区的 100 积分。积分可用于在异步社区兑换优惠券、样书或奖品。



扫码关注本书

扫描下方二维码，您将会在异步社区微信服务号中看到本书信息及相关的服务提示。

与我们联系

我们的联系邮箱是 contact@epubit.com.cn。

如果您对本书有任何疑问或建议，请您发邮件给我们，并请在邮件标题中注明本书书名，以便我们更高效地做出反馈。

如果您有兴趣出版图书、录制教学视频，或者参与图书翻译、技术审校等工作，可以发邮件给我们；有意出版图书的作者也可以到异步社区在线提交投稿（直接访问 www.epubit.com/selfpublish/submission 即可）。

如果您是学校、培训机构或企业，想批量购买本书或异步社区出版的其他图书，也可以发邮件给我们。

如果您在网上发现有针对异步社区出品图书的各种形式的盗版行为，包括对图书全部或部分内容的非授权传播，请您将怀疑有侵权行为的链接发邮件给我们。您的这一举动是对作者权益的保护，也是我们持续为您提供有价值的内容的动力之源。

关于异步社区和异步图书

“**异步社区**”是人民邮电出版社旗下 IT 专业图书社区，致力于出版精品 IT 技术图书和相关学习产品，为作译者提供优质出版服务。异步社区创办于 2015 年 8 月，提供大量精品 IT 技术图书和电子书，以及高品质技术文章和视频课程。更多详情请访问异步社区官网 https://www.epubit.com。

“**异步图书**”是由异步社区编辑团队策划出版的精品 IT 专业图书的品牌，依托于人民邮电出版社近 30 年的计算机图书出版积累和专业编辑团队，相关图书在封面上印有异步图书的 LOGO。异步图书的出版领域包括软件开发、大数据、AI、测试、前端、网络技术等。

异步社区

微信服务号

目录

第 1 章

域间路由和 Internet 组播

本章将介绍域间组播流量转发的基本要求，以及域间组播设计的三大基本要素：用来标识组播源主机的控制平面、用来标识组播接收主机的控制平面，以及下游控制平面。

1.1 域间组播简介

很多应用程序可能需要在形形色色的大型网络（如 Internet）内得到组播的支持。组播发送主机可能位于某种网络之内，而潜在的组播接收主机可能位于另一种网络之内。组播接收主机和组播发送主机可存在于完全不同的网络之内，两种主机所在的网络可分别由不同的管理机构来管控，管理机构会各自采用不同的策略来掌控组播流量的转发。

在上述情况下，不能指望组播流量转发路径沿途的所有 L3 设备共享同样的配置或策略。这就是基于 Internet 的组播应用程序所面对的现实。因此，为了在不同的网络之间提供组播服务，除了基本的组播传输协议之外，还得仰仗其他协议，需要额外的配置。为什么会这样呢？莫非大型互联网络（比如，Internet）要用另外一种 Internet 协议（IP）来传递组播流量？要是每个网络都分别根据不同的规则来进行管理，那么 Internet 又是怎么把这些网络互连在一起的呢？如本书第 1 卷所述，协议无关组播（PIM）协议是 IP 组播事实上的标准转发协议。既然 PIM 是通用协议，那为什么需要制定不同的策略呢？对 Internet 原理的简单介绍不但可以回答上述问题，还能帮助读者理解为什么在设计跨网通信的组播应用程序时需要额外的考量。

Internet 协议在发明时就被视为一种“尽力而为”（best effort）服务。即便将 IP 推广至万维网（WWW），能用其互连多个网络，尽力而为也是通用准则。这条准则至今依然成立，它不但决定了当今 Internet 的转发行为，同样决定了其他任何一种大型多区域（multidomain）网络的转发行为。在这样的环境里，“尽力而为”是指当 IP 流量在网络之间传递时，流量转发路径沿途的每个穿越网络（transit network）都得到了适当地配置，目的是力争以最好的方式将流量转发至正确的目的网络。但是，既不保证流量会以最优方式转发，也不确保流量能

抵达最终的目的网络。

注意 尽力而为的转发模式在概念上是统一的，适用于所有 IP 流量：单播、组播和广播流量。本节简要探讨了尽力而为的基本概念，进一步说明了在不同的网络之间传递组播流量时，为什么需要开启额外的转发机制。虽然本节介绍的都是基本概念，但为理解组播流量的跨网转发奠定了基础。本章假定读者对单播 Internet 转发有着清楚的认识，包括边界网关协议（BGP）的用法。

要是从宏观的角度来审视 Internet，就会发现它实际上是一张大网，由互连在一起的各色各样的网络构成。通常，Internet 服务提供商（ISP）之间会彼此互连，目的是与别的服务提供商或客户建立对等关系，实现互连互通，传递穿越流量。

ISP 连接末端客户（比如，家庭用户、蜂窝网络、中小企业、研究机构、医院、政府、企业等）的方式多种多样。图 1-1 以一家虚构的名为 Mcast Enterprises 的公司为例，来描述 ISP 之间以及 ISP 和客户之间的互连方式。

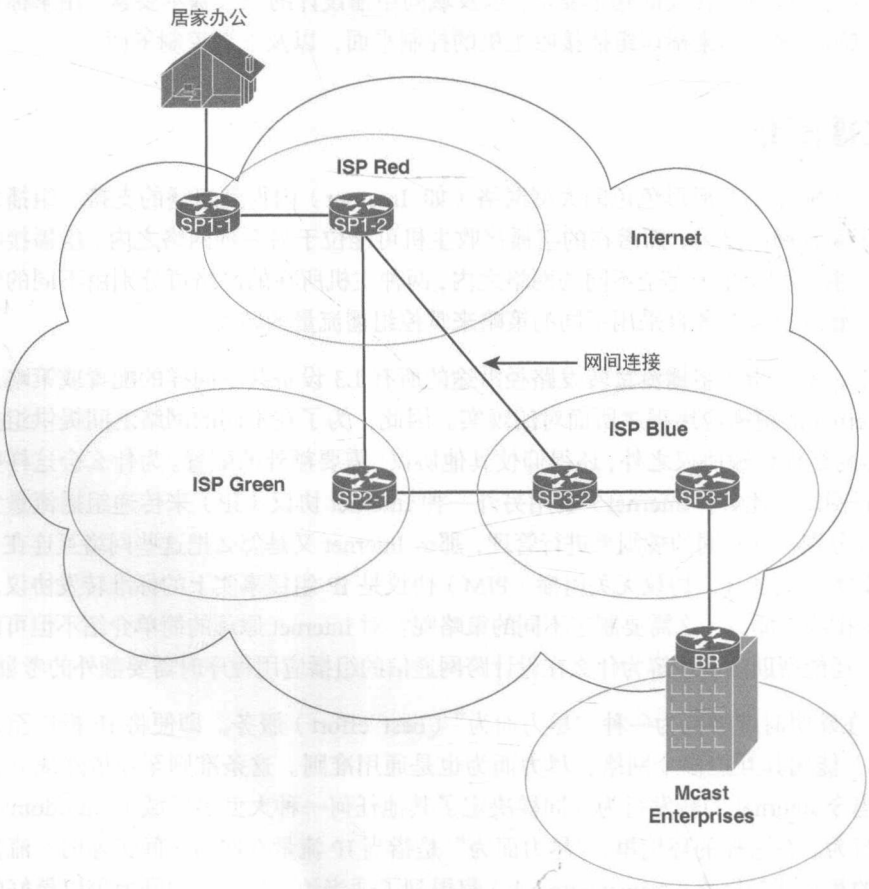


图 1-1 Mcast Enterprises 网络与 ISP 网络的连接示意图

出现在图 1-1 中的每一个网络实体（用椭圆包围）都称为自治系统（AS）——一个具有管理和运维边界的网络，其自身与任何其他 AS 之间会划定明确的边界。与 IP 地址一样，自治系统也用编号来标识，编号的分配受控于 Internet 编号分配机构（IANA）。图 1-2 所示为前图以 Internet AS 方式呈现的示例网络，其中的 AS 用简单的圆圈来表示，采用的是私有自治系统编号（Autonomous System Number，ASN）。

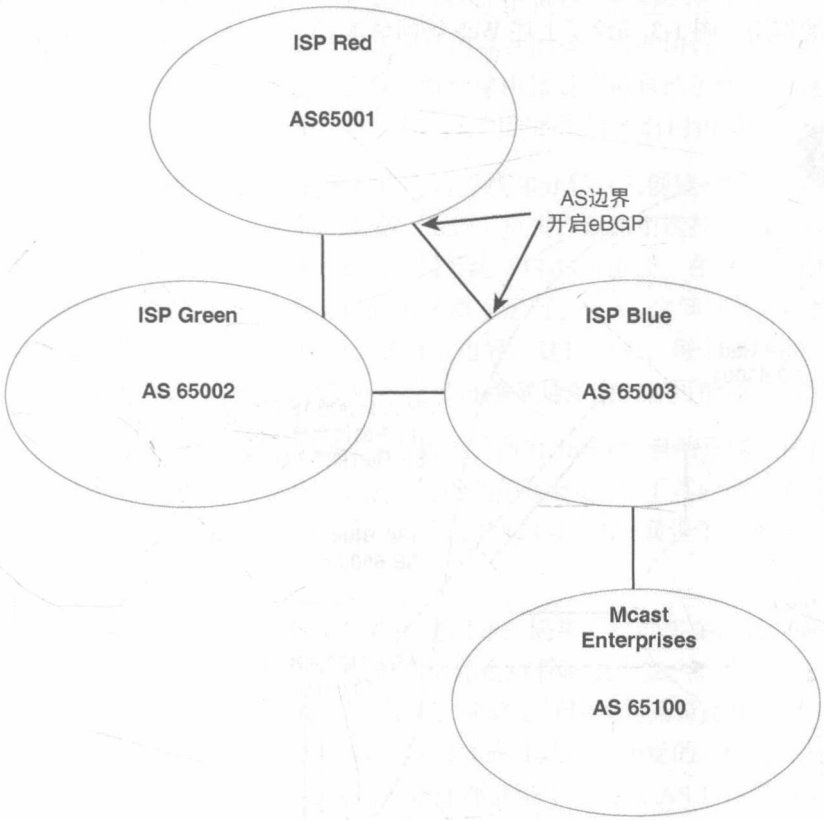


图 1-2 Mcast Enterprises 网络，相互连接的 AS 中的一员

注意 根据 IETF 的定义，ASN 跟 IP 地址一样都是公用的。不过，ASN 也有一个专用号段被预留供非公用网络使用，这也跟 IP 地址一样。标准的 16 位私用 ASN 号段为 64512～65535，由 RFC 6996（2013 年从 RFC 1930 更新而来）定义。尽管本书可能会探讨各种 Internet 功能，但采用的所有编号（IP 编址、ASN 和组播组地址编号）都是私用的，目的是不与现有的 Internet 服务相混淆，保护公众的利益。

互连的 AS 之间会通过共享某些路由信息，来拼凑完整的 Internet 拓扑。尽力而为的转发模式意味着，路由器查询目的网络信息，流量在 AS 之间穿行，每个 AS 自有一套转发规则。