经典译丛

信息网络技术与网络科学

The Illustrated Network How TCP/IP Works in a Modern Network

现代TCP/IP网络详解

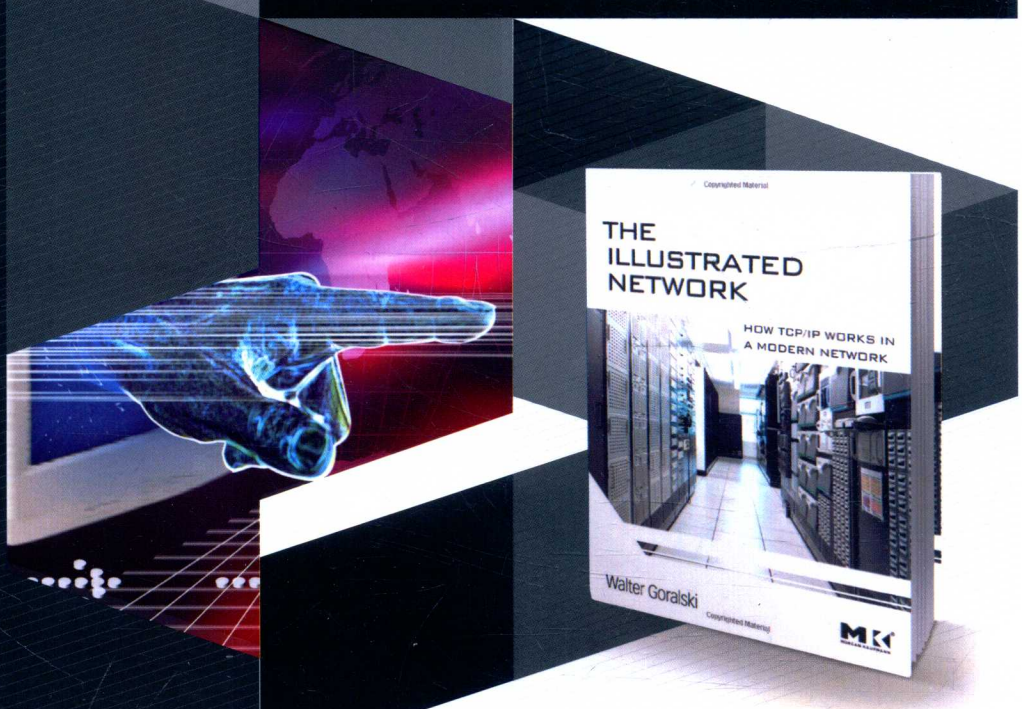
The Illustrated Network

How TCP/IP Works in a Modern Network

【美】 Walter Goralski 著

黄小红 闫 岫 张 沛 赵 钦 等译

马 严 审校



ELSEVIER
爱思唯尔

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

经典译丛·信息网络技术与网络科学

现代 TCP/IP 网络详解

The Illustrated Network

How TCP/IP Works in a Modern Network

[美] Walter Goralski 著

黄小红 闫 岫 张 沛 赵 钦 等译

马 严 审校

電子工業出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书讲解各种网络协议如何在更大、更复杂的现代网络中实际应用并不断发展。综合运用自20世纪90年代中期以来的所有网络新技术构建了一个复杂的网络，通过使用服务器和路由器等设备，以一个网络为例详细解释了TCP/IP协议如何运行。示例网络与真实运转网络的配置是一致的，包括服务器、路由器和工作站，其中每个对象都是来自于真实运行在网络中的互联实体。

本书适合作为高等院校计算机网络及相关专业课程的教材，也适合网络管理员、运维人员、工程师、系统设计师和管理者以及其他需要了解网络知识的技术人员（如芯片设计者、电信工程师、软件工程师等）参考。

The Illustrated Network: How TCP/IP Works in a Modern Network
Walter Goralski
ISBN: 9780123745415

版权贸易合同登记号 图字：01-2010-4323

图书在版编目(CIP)数据

现代TCP/IP网络详解/(美)戈拉尔斯基(Goralski, W.)著；黄小红等译. —北京：电子工业出版社，2015.10
(经典译丛·信息网络技术与网络科学)
书名原文：The Illustrated Network: How TCP/IP Works in a Modern Network
ISBN 978-7-121-26483-2

I.①现… II.①戈… ②黄 … III. ①计算机网络-通信协议 IV. ①TN915.04

中国版本图书馆CIP数据核字(2015)第143938号

策划编辑：马 岚
责任编辑：马 岚　　特约编辑：姚 旭
印　　刷：北京京师印务有限公司
装　　订：北京京师印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路173信箱　邮编　100036
开　　本：787×1092　1/16　印张：34.25　字数：943千字
版　　次：2015年10月第1版
印　　次：2015年10月第1次印刷
定　　价：99.00元

凡所购买电子工业出版社的图书有缺损问题，请向购买书店调换；若书店售缺，请与本社发行部联系。联系及邮购电话：(010)88254888。

质量投诉请发邮件至zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

服务热线：(010)88258888。

译 者 序

学习计算机网络技术，特别需要理论与实践相结合。在各种 TCP/IP 的论著中，能将理论与实践相结合的，最著名的就是 W. Richard Stevens 的《TCP/IP 详解》。这套书共三卷，结合网络配置、数据截图等方式讲解了网络互联的基本概念和机制，网络协议的细节和实现，以及应用层的设计原则和技术。对于从事计算机网络的教师以及科研人员，这套书是必读的。

本书则是对《TCP/IP 详解》的改进。作者通过与 Juniper 网络公司的合作，搭建了一个真实的、接近生产规模级的网络，该网络包含了通过 SONET 链路连接的六个路由器、两个以太局域网、两对 Windows 客户端和服务器、一对 Linux 主机和一对 FreeBSD 主机，这是一般计算机网络科研人员难以做到的。本书中每章的内容都基于同样的网络拓扑和 IP 地址进行阐述，随着技术的不断深入，逐渐完成对路由器、客户端以及服务器的配置，生动、灵活地给读者展示了每种技术在实际网络中的效果，这是纯粹的理论知识介绍的书籍所无法包含的。

本书涵盖了 TCP/IP 网络中的大部分热门技术，与经典的 TCP/IP 书籍一样，采用了由下而上的方法介绍协议栈，概念清楚、准确，讲解详细，例子很多。作者在书中所介绍的协议或技术均在作者搭建的真实网络中进行了实验验证，详细给出了网络配置过程以及数据交互的截图，并针对每章的内容留有一定数量的思考题。

对于学习和了解计算机网络的本科生和研究生，本书可以作为他们的教材和教学参考书。对于那些从事网络工程和技术服务的技术人员，如网络管理员、运维人员、工程师、系统设计师和管理者，以及其他需要了解网络知识的技术人员（如芯片设计者、电信工程师、软件工程师等），本书也是一本很好的理论和实践相结合的参考读物。

参加本书翻译的有：黄小红、闫岫、张沛、赵钦、苑婷婷、牟澄、卢军强。全书由马严教授审校。

我们曾经期望按照中文习惯来叙述本书的内容，但是限于精力和水平，最后未能完全做到这一点，翻译中也难免有不妥或错误之处，敬请读者谅解和指正。

序

世纪交替之际，网络融合已经成为行业发展趋势。通过将数据、语音、视频、虚拟专用网(VPN)以及其他服务融合到一套共享的基础设施中，除了能够减少投资，带来巨大的经济利益以外，更能够降低运维成本，使得互联网提供商不必针对某种服务而运维特定的设备和系统。网络融合的基本原则是使用一套相同的基础设施支持多样化的服务，即通过一个通用的封装协议，满足不同服务的传输需求——互联网协议(Internet Protocol)就是这样的协议。

一切基于 IP

网络行业发展日新月异。10 年乃至更短的时间就足可以使一个新兴的前沿技术由萌发走向凋亡(如 ATM、帧中继、令牌环或 FDDI 等)。然而，自 20 世纪 60 年代早期有了最初的想法，TCP/IP 奇迹般地走过了 35 年的历史，并逐步发展壮大，至今仍在不断演进、前行。

这个由 Vint Cerf 与 Bob Kahn 在 1973 年提出的协议已经经历了(且将继续经历)数百次的改进和一次版本的重大升级，然而其核心功能与之在 20 世纪 80 年代中期时的状态几乎无异。在一个优胜劣汰极为残酷的技术行业里，TCP/IP 的高龄强有力地证明了其过人之处。

目前为止，并没有任何明显迹象表明 IP 技术将要退出历史舞台。与此相反，越来越多基于 IP 的新应用、新设备和新服务不断涌现，而 IP 协议的新版本——IPv6 协议，因其可以为未来网络发展提供足够的地址空间，网络升级过渡的必要性也日渐凸显。虽然写本序时 IPv6 尚处于部署的早期阶段，但可能读者将来某一天阅读本书时，IPv6 已成为大多数人所知道的唯一 IP 版本。

TCP/IP 逐步统治网络行业的故事早已在业内广为流传。Cerf、kahn、Jon Postel 等人以及其他对 TCP/IP 早期发展做出贡献的人同样是 ARPANET(现代互联网的前身)的创始成员。在协议栈被集成到 UNIX 系统后，TCP/IP 受到了开发者的追捧和欢迎，这为之奠定了雄厚的基础，使得其有更多的机会扎根于早期尚处在襁褓阶段的网络工业中。

早期，TCP/IP 被认可的程度并不理想。以政府部门和电信公司为代表的组织机构对于工作组那套“先让它试试，看看到底能怎样”的不正式的标准化过程感到难以接受，这些严谨的组织更愿意使用那些通过严格标准过程所产出的协议。毕竟在当时谁也无法相信，一群满腔热血的学生竟然最终促成了如今的互联网工程任务组(IETF)。因此，国际标准化组织(ISO)被任命来开发一个“成熟的”网络协议族，这一协议族最终发展成为后来的开放系统互连(OSI)。

ISO 擅长设计复杂而详尽的技术标准，并且只有当标准处于完成、可生产的状态下才进行发布，这种做法需要消耗大量的时间。OSI 主张先使用 TCP/IP 作为临时的解决方案，待到 ISO 标准委员会完成全部工作之后再做打算。然而，等 OSI 做好准备时，TCP/IP 早已被广泛部署、验证和熟知，极少有网络运营商愿意将业务迁移到新协议上。

OSI 之所以保留到今天，主要是因为 IS-IS 协议和那无所不在的 OSI 参考模型，而 TCP/IP 协议则成为了全球性的通信传输协议标准。

图解网络

我是一个注重图形表达的人。我欣赏很多同事通过口头详述即可构建网络场景的能力，而我则更擅长使用图形的方式来完成对其的阐述与理解。

1994 年，W. Richard Stevens 发布了《TCP/IP 详解》的第一卷，它马上成为了我最喜欢的书籍之一，同时也排在了我给学生推荐的参考书目清单的首位。Stevens 使用图表、配置以及数据截图来讲授 TCP/IP 协议栈，这使得该书不仅仅是一本成功的教科书，更是一本全面的案例研究。本书的可视化效果非常好，读者完全不必坐在协议分析器前就能观察数据包的来回传输。

尽管 Stevens 的著作在阐释单独的 TCP/IP 组件的行为方面已经非常出色，但它并没有诠释这些组件在一个大型的真实网络中是如何交互的。

Walt Goralski 着重对此做出了改进。本书同样使用了由下而上的方法（Stevens 的原话）来讲授协议栈：每一章都建立在前面章节的基础上，向读者展示协议是如何工作的。通过运行 Juniper 网络，读者不仅能够看到实验室级别的设备交互，还能看到当今生产级规模的网络场景，进而领略作者在真实、可视的自然状态下对 TCP/IP 协议的探索与研究。

本书必将成为 IP 网络实际应用的经典书籍之一，会是学生必读书目列表的基石之一，受到专业人士的喜爱。

Jeff Doyle
Westminster，Colorado

前　言

这并不是一本教你如何使用互联网的著作，而是告诉读者互联网是如何成为对用户有用的网络的。互联网是一个公共的、全球性的网络，基于 TCP/IP 协议，故其通常又称为互联网协议族。网络协议就是一系列为了完成某些功能所必须遵守的规则。TCP/IP 实际上是传输控制协议(TCP)和互联网协议(IP)两个协议的合成，这两个协议可以使信息在网络之间进行传输，是互联网早期最先使用的两个协议。TCP/IP 是现代网络的核心与灵魂，本书则使用图解的方式介绍了它们的工作流程。通过跟踪数据在各种互联网连接中的传输过程，读者可以观察到现今的网络是如何在 TCP/IP 的支撑下完成工作的。

读者对象

本书主要介绍通用网络技术，尤其是互联网技术。我并不肯定丝毫网络知识经验都没有的人可以理解整本书籍的内容。但是，任何只要在线收发过邮件、浏览过网页、下载过电影或歌曲，与世界各地的亲友聊过天的人，在理解本书内容时都不会感到过于困难。

在每章的末尾都留有思考题，这使得本书可以作为高中学生或者计算机科学和电子工程院系的大学生研修第一门计算机网络课程的教科书。同时，它又不仅仅是一本教科书，对于那些初涉电信行业或致力于入职以互联网为其业务计划重要组成部分的公司(这样的公司越来越多)的读者而言，本书具有同样的宝贵价值。

书中只有一章用到了 C 语言的代码，并且它仅仅是为了给读者提供一些必要信息，所有高中阶段没有纳入的数学内容在本书中都没有被提及。书中没有微积分、没有概率论，没有随机过程等等内容，一切内容简单易懂。一些必须用到数学功底的内容，比如公钥加密内容中使用的“便携式计算机”以及 Diffie-Hellman 密钥分配案例等，作者都认真地进行了设计，使得数学内容变得尽可能简单。

本书独一无二的特点

本书的过人之处在于你很难在其他同类书籍中找到有关 TCP/IP 的如下内容：

1. 书中每一章的每个案例都基于相同定义的网络拓扑和网络地址；
2. 同等对待 IPv4 和 IPv6 协议，同步介绍两种互联网协议涉及的所有概念；
3. 除 TCP/IP 协议外，本书还讲解了路由协议以及 TCP/IP 的上层应用；
4. 书中讨论的网络模型基于真实的网络运营商和企业级的以太网；
5. 既讲解利用协议所能实现的服务和功能，又讲解功能背后的协议原理；
6. 涵盖了诸如 MPLS、IPSec 等一般讲解 TCP/IP 书籍中所不涉及的话题。

那么，本书采用此种方式阐述的初衷为何呢？我观察到，即使是生活在如今互联网发达的时代里，仍然很少有人研究整个网络，更鲜有将路由器、TCP/IP、互联网和大量相关主题融会

贯通的案例。看上去讲了不少，学了很多，但是当考虑与这些议题相关的复杂原理时你就会发现，泛在的计算机“扫盲教学”和基本编程语言课程真正所覆盖的内容少之又少。

当我写这本书并将它打印出来放在我工作的地方时，一名芯片设计师发现了它，并饶有兴趣地开始阅读。当我回来取打印稿时，他正着迷于某些章节，并当场表示对于本书的获取意愿。通过与他的交谈，我意识到成千上万的人进入了互联网行业，其中很多人都来自其他学科和领域。随着互联网的发展，社会对于数字通信结构的依赖仍在继续，越来越多的人需要这样的综述来了解现今的网络是如何运转的。仅靠本书中零星的网络知识肯定无法满足读者的求知欲和好奇心。我希望每一位读者都能寻找到一种方法，增长对于某个感兴趣领域的知识。这本书包含了数百个网络议题，而每个议题都是错综复杂的，全都详加阐述则需要很多篇幅。举个例子，有二三十本书都描述了 MPLS 的原理及其演进技术，而本书的相关章节却只有二三十页。希望本书能够通过“图解”的方式来介绍现代网络的工作方式，并能帮助想了解 MPLS 的人找到那二三十本书，进而了解网络整体的面貌和运行原理。

像很多人一样，我学习网络知识，包括路由器和 TCP/IP，主要是来源于书籍和别人口传面授，其中所缺少的内容则通过网络实践来弥补。我所看过的书籍都非常好，介绍了这项或那项技术是如何工作的，但是我常常无法看到真正工作时的场景。这本书则是知行合一，尝试给读者一个机会去观察真实的、运行中的、具有合适大小和规模的、可实际操作的网络环境，去了解各项任务是如何完成的，以探究屏幕背后的工作原理。毕竟，了解访问网站时发生了什么是一码事，而看到其背后具体的动作是另一码事。

本书的目的是让读者看到当其访问网站、写邮件、下载歌曲或通过互联网在电话中交谈时，网络中究竟发生了什么，使读者通过观察而了解现代网络的工作方式。

本书中学不到什么

列出书中不会涉及的内容看上去挺奇怪，但这总比读者通读全书后才发现并没有想找的东西要好。下面列出了本书不包括的内容。

本书并没有提到广受关注与使用的点对点协议（用于在全网分布式地部署服务器），没有提及聊天室或聊天服务使用的协议，也没有探索音乐和视频的下载服务。换句话说，书中不包括 YouTube，IRC，iTunes 或 eBay 的内容。

这些议题当然很有趣，也很重要，但限于篇幅，我们只能专注最重要的部分。

图解网络

在学习过程中，过于简单的实验性网络设置往往会阻挠读者进一步深入了解知识本质，他们渴望一个更复杂的、真实的、安全的网络环境，进而深入地探索 TCP/IP 协议、分层和应用，而不用担心他们所看到的仅仅局限于一个实验室环境，或担心某些实验可能导致整个网络瘫痪。

整个网络的客户端、服务器和路由器被少部分人在晚上或者周末时间整体操控的时代早已过去。现在的网络无时无刻不在工作，尤其是现在这个时代，在地球上的一半用户正在睡觉时，地球上的另一半用户可能处于上网的“黄金时间”。

过去，在我遇到新功能或程序并且告诉自己：“但愿有机会上手试试，看看它到底会有怎样的效果”时，会有各种阻力妨碍我。如今，在经历了近 40 年的网络实践工作之后，我才刚到

达了终于可以坦言“我想去试试……”的境界，而此时也终于没人来告诉我“这个不能做”了。

我所就职的公司 Juniper Networks Inc. 鼎力支持了这本书籍的撰写，使得我不仅可以讨论 TCP/IP 协议或协议运行的案例，更能通过一系列客户端、服务器、路由器和连接（包括公共互联网）来图解网络中每一部分的深层原理。公司拥有足够的路由器和链路，并提供了所有可能需要的 UNIX 和 Windows 主机（反思之，事实上这些设备对于介绍网络来讲略有冗余，在大部分的章节里只用到少数路由器而已）。为了便于操作，我们没有把 XP 的主机升级到 Vista（当时 Vista 刚刚出现不久），同时保留了当时比较流行的 IE6。

基于 Juniper Networks Inc. 的帮助，我们构建了本书中需要的网络。虽然它花费了一点额外的时间，但物有所值。网络里包含了通过 SONET 链路连接的六个路由器、两个以太局域网、两对 Windows XP 的客户端和服务器（家庭和专业版）、一对红帽 Linux 主机（RH 9 内核 2.4.20-8）以及一对 FreeBSD（4.10 版）的主机。

图 P.1 展示了我们搭建的网络，该网络将出现在本书的每一章，用图解方式来说明所讨论的网络概念。

阅读方式

本书要求从开始到结束逐章按顺序进行阅读。这样的要求似乎很有趣，因为现在很多技术类书籍并不需要像读小说或传记那样的方式进行阅读。读者更倾向于在这些书中去找寻特定的目标，然后从特定的地方开始阅读。当然，这是读者的自由，但我们仍然建议采用逐章阅读的顺序。

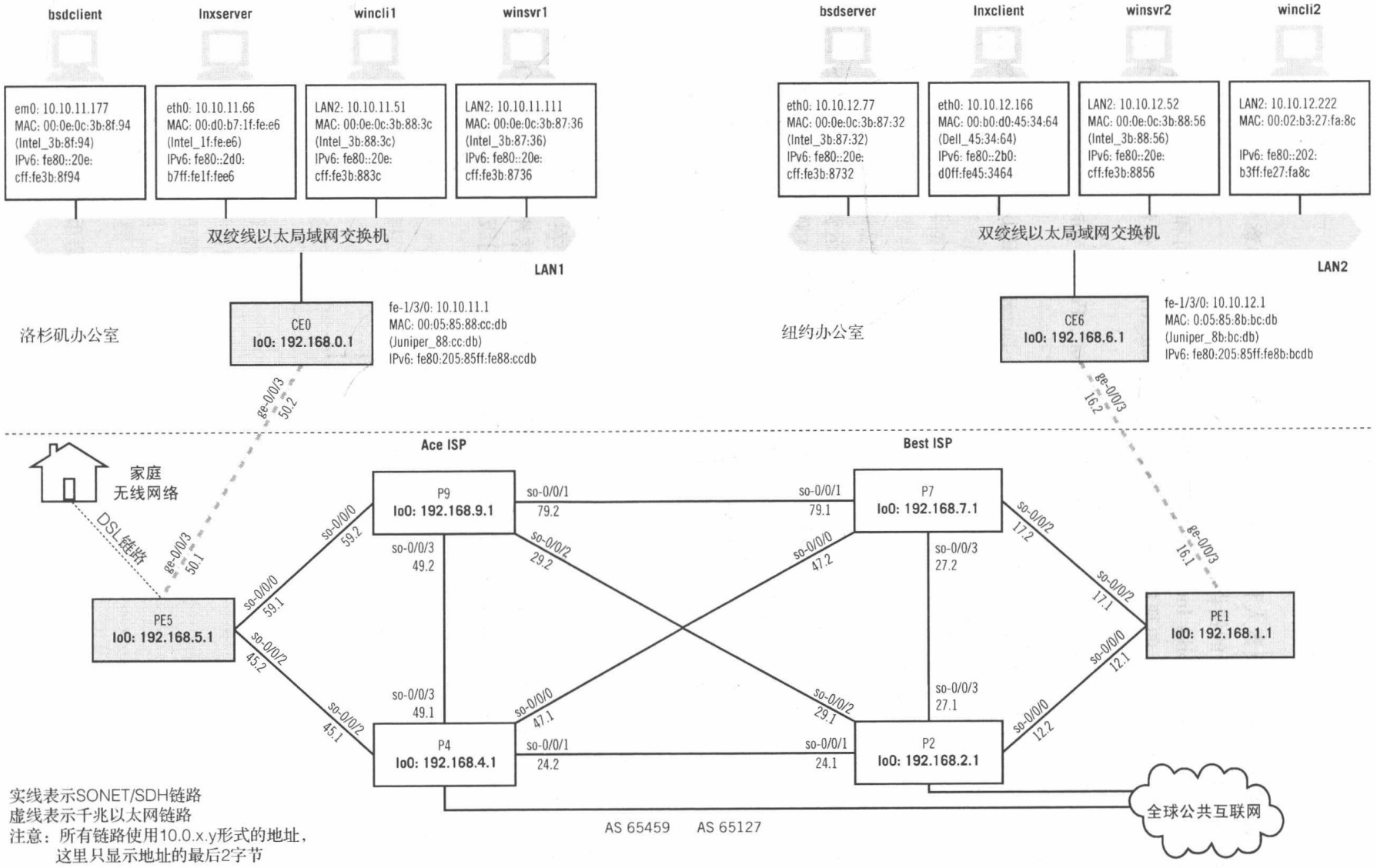
如果不能像冒险故事那样扣人心弦，我仍然希望书中的故事如解谜一般环环紧扣。本书首先在第 1 章为读者提供了分层协议的介绍，之后直至结尾，讲述了一个以逻辑方式组织的自下而上贯穿整个互联网协议族的精彩故事。所以，如果时间足够，请你从头读到尾，因为每章的内容都依赖于前一章的知识基础。如果对网络概念还不了解，或刚刚开始了解，建议采用这种连续的方法阅读。对于那些职场老手，当然用哪种方式阅读都可以，但请记住所有的重点在本书中都是同等重要的。从中间开阅读时，读者可能会对某个议题所包含的内容产生疑惑，此时则需要从前面的章节中寻找答案。

开始阅读时通常会发现，每一章的结构都基本相同。前面的章节可能会和后面的章节格式不同，因为前面的章节要求对协议、应用或概念进行探索。而后面的章节则逐步开始图解和剖析协议是如何工作的。在某些情况下，书中讨论的内容不仅包含我们所搭建的网络，还将涉及整个全球互联网。需要注意的是，每章中网络配置的细节或多或少会有所变化，特别是那些涉及路由器的配置，这些变化在发生时会有详细的说明。

本书配套网站是 www.elsevierdirect.com/companions/9780123745415，从中可以找到书中采集的许多文件，从而能够自行探索相关协议。

源代码

第 3 章讲解网络技术时使用了 Aeropeek 提供的无线网络采集的案例。第 12 章讲解套接字时使用了 Michael J. Donahoo 和 Kenneth. L. Calerts 所著的 *TCP/IP Sockets in C*（Morgan Kaufmann，2001）中的案例程序。感谢这两个团队允许我使用其资料。


bsdclient
lnxserver
wincli1
winsvr1
em0: 10.10.11.177
MAC: 00:0e:0c:3b:8f:94
(Intel_3b:8f:94)
IPv6: fe80::20e:
cff:fe3b:8f94
eth0: 10.10.11.66
MAC: 00:d0:b7:1f:fe:e6
(Intel_1f:fe:e6)
IPv6: fe80::2d0:
b7ff:fe1f:fee6
LAN2: 10.10.11.51
MAC: 00:0e:0c:3b:88:3c
(Intel_3b:88:3c)
IPv6: fe80::20e:
cff:fe3b:883c
LAN2: 10.10.11.111
MAC: 00:0e:0c:3b:87:36
(Intel_3b:87:36)
IPv6: fe80::20e:
cff:fe3b:8736
双绞线以太局域网交换机
LAN1
洛杉矶办公室
CE0
lo0: 192.168.0.1
fe-1/3/0: 10.10.11.1
MAC: 00:05:85:88:cc:db
(Juniper_88:cc:db)
IPv6: fe80:205:85ff:fe88:ccdb
ge-0/0/3
50.2
bsdserver
lnxclient
winsvr2
wincli2
eth0: 10.10.12.77
MAC: 00:0e:0c:3b:87:32
(Intel_3b:87:32)
IPv6: fe80::20e:
cff:fe3b:8732
eth0: 10.10.12.166
MAC: 00:b0:d0:45:34:64
(Dell_45:34:64)
IPv6: fe80::2b0:
d0ff:fe45:3464
LAN2: 10.10.12.52
MAC: 00:0e:0c:3b:88:56
(Intel_3b:88:56)
IPv6: fe80::20e:
cff:fe3b:8856
LAN2: 10.10.12.222
MAC: 00:02:b3:27:fa:8c
IPv6: fe80::202:
b3ff:fe27:fa8c
双绞线以太局域网交换机
LAN2
纽约办公室
CE6
lo0: 192.168.6.1
fe-1/3/0: 10.10.12.1
MAC: 0:05:85:8b:bc:db
(Juniper_8b:bc:db)
IPv6: fe80:205:85ff:fe8b:bcdb
ge-0/0/3
16.2
家庭
无线网络
DSL链路
ge-0/0/3
50.1
Ace ISP
Best ISP
PE5
lo0: 192.168.5.1
P9
lo0: 192.168.9.1
P4
lo0: 192.168.4.1
P7
lo0: 192.168.7.1
P2
lo0: 192.168.2.1
PE1
lo0: 192.168.1.1
so-0/0/0
59.1
so-0/0/2
45.2
so-0/0/0
59.2
so-0/0/3
49.2
so-0/0/3
49.1
so-0/0/2
45.1
so-0/0/1
79.2
so-0/0/2
29.2
so-0/0/0
47.1
so-0/0/2
29.1
so-0/0/1
24.2
so-0/0/1
79.1
so-0/0/0
47.2
so-0/0/3
27.2
so-0/0/3
27.1
so-0/0/1
24.1
so-0/0/2
17.2
so-0/0/2
17.1
so-0/0/0
12.2
so-0/0/0
12.1
ge-0/0/3
16.1
全球公共互联网
AS 65459
AS 65127
实线表示SONET/SDH链路
虚线表示千兆以太网链路
注意：所有链路使用10.0.x.y形式的地址，
这里只显示地址的最后2字节

图P.1 示例网络

致谢

我要感谢每个领域的负责人利用他们的时间阅读和审查选定的章节。他们的建议使这本书变得更好，书中任何遗留的错误都算我的。

我要感谢 Juniper Networks Inc. 的同事们，他们耗费大量的时间和精力搭建了这个网络，他们也为这本书提供了巨大的帮助。首先要感谢的是 Scott Kriens，他帮忙构建了一个适合创新和探索的网络环境。谢谢你，Scott!

其次还要感谢 June Loy，Aviva Garrent，Michael Tallon，Partick Ames，Jason Lloyed，Mark Whittiker，Kent Ketell 以及 Jeremy Pruitt。

最后，我还想要感谢本书的主要技术评阅专家，Joel Jaeggli 和 Robin Primentel，感谢他们的认真审查、仔细修改并提出了很多建议。

主要技术评阅专家

Joel Jaeggli 在诺基亚安全和移动连接小组工作。他的工作包括运营 nokia. net(AS 14277)研究网络和支持诺基亚安全业务的战略规划。他之前在俄勒冈大学工作，参与了很多项目，包括网络创业资源中心、俄勒冈州路由视图项目、Beyond BGP 项目和俄勒冈州的视频实验室。他活跃于多个工业界相关团体，包括 IETF(工作组主席)和 NANOG 项目(两届程序委员会成员)。Joel 经常作为一个指导者或者讲演者参与以互联网服务和安全为主题的地区性及国际网络会议。

Robin Primental 目前是 Facebook 的网络工程师，负责维护网络，使网络可随 Facebook 的用户和应用程序的增长而增长。此前，Robin 在谷歌和雅虎的产品网络团队工作。Robin 也花了 6 年在 Teradyne 担任过网络、安全和 UNIX 底层结构工程师的工作。在他开始其计算机网络职业生涯之前，Robin 曾在益华电脑系统设计和英特尔公司工作。在芯片部门工作时，Robin 专门从事硅芯片布局和路由、VHDL 行为逻辑验证以及片上存储器的门级逻辑验证。

目　录

第一部分　网络基础

第二部分　核心协议

第三部分 路由和路由协议