经典译丛

信息网络技术与网络科学

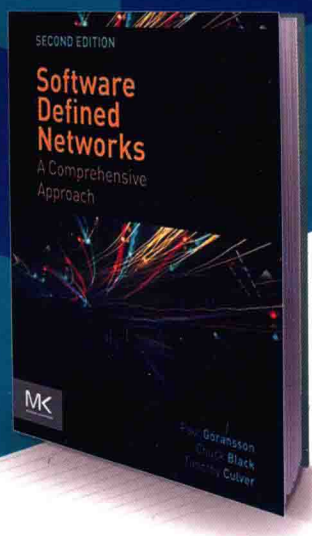
深度剖析软件定义网络（SDN）（第二版）

Software Defined Networks
A Comprehensive Approach, Second Edition

【美】 Paul Göransson
Chuck Black
Timothy Culver 著

王 海 张 娟 等译

Software Defined Networks: A Comprehensive Approach, Second Edition



ELSEVIER

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

深度剖析软件定义网络（SDN）

（第二版）

Software Defined Networks

A Comprehensive Approach

Second Edition

Paul Göransson

[美] Chuck Black 著

Timothy Culver

王 海 张 娟 郭 晓 李艾静

王向东 于卫波 米志超 朱 毅 译

陈 娟 徐正芹

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

内 容 简 介

软件定义网络（SDN）是一种新型网络架构，旨在帮助网络适应快速变化的业务需求。本书全面介绍了 SDN 的基本概念、原理和商业应用。全书共 15 章，从多个方面阐释了 SDN，从 SDN 的由来和历史沿革，到 SDN 技术的最新发展情况和未来走向，全面分析了 SDN 和 OpenFlow 的技术原理，分析了 SDN 的开源代码及相关资源，并探讨了 SDN 的应用场景、商业发展及其局限性，是一本视野广阔的 SDN 参考手册。第二版新增了网络功能虚拟化（NFV）以及新兴的协议、控制器和应用模型两章，并更新了部分技术知识。

本书对从事 SDN 技术研发的专业人士、设备商、云服务提供商、数据中心管理人员、企业 IT 运维人员、科研工作者等多个领域的从业人员具有参考价值，也适合作为高等院校计算机、网络、通信等专业的 SDN 选修课教材。

Software Defined Networks: A Comprehensive Approach, Second Edition
Paul Göransson, Chuck Black, Timothy Culver
ISBN: 9780128045558

版权贸易合同登记号　图字：01-2018-0204

图书在版编目（CIP）数据

深度剖析软件定义网络：SDN：第二版/（美）保罗·戈朗生等著；王海等译. —北京：电子工业出版社，2019.7
（经典译丛. 信息网络技术与网络科学）
书名原文：Software Defined Networks: A Comprehensive Approach, Second Edition
ISBN 978-7-121-36759-5

①深…　II. ①保…　②王…　III. ①计算机网络—研究　IV. ①TP393

中国版本图书馆 CIP 数据核字（2019）第 106585 号

责任编辑：杨　博
印　　刷：三河市君旺印务有限公司
装　　订：三河市君旺印务有限公司
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：787×1 092　1/16　　印张：22　字数：563 千字
版　　次：2016 年 3 月第 1 版（原著第 1 版）
　　2019 年 7 月第 2 版（原著第 2 版）
印　　次：2019 年 7 月第 1 次印刷
定　　价：109.00 元

凡所购买电子工业出版社图书有缺损问题，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）88254888，88258888。

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。

本书咨询联系方式：yangbo2@phei.com.cn。

译 者 序

我们非常高兴有机会向广大读者推荐本书。

软件定义网络是近年来网络发展的一个重点方向，它与大数据、云计算一起吸引了众多科技工作者、管理者和广大用户的注意力。铺天盖地的宣传与产品已经在市场上涌现，但对于大多数读者来说，存在“到底什么是软件定义网络？”“软件定义网络到底好在哪里？”“软件定义网络是否适合我？”的疑问。本书以通俗易懂的语言、涵盖广泛的图表，全面、深入、准确地解答了上述问题。

今天，在互联网上充斥着各种有关软件定义网络的信息，包括各种形式的博客、播客、微博，等等。然而，正如读者将在本书里看到的那样，软件定义网络这个名词下面实际上囊括了多种完全不同的技术、体制以及具体的实现系统，大部分针对软件定义网络的介绍都只是讨论了这些技术的一部分，并且由于观察角度和观察对象的不同，难免以偏概全，这也是很多专业人士对软件定义网络的观点大相径庭的原因。市面上关于软件定义网络的书也有不少，但从整体看，有的书侧重描述协议细节，对于技术的整体缺乏阐述，只见树木不见森林；有的又过于宏观，缺乏细节。本书作者 Paul Göransson 博士和 Chuck Black 先生凭借他们多年浸淫在软件定义网络里的工程和实践经验，站在相对公正的，或者说更贴近 SDN 起源意图的立场上，采用先整体、后局部，先宏观、后微观的阐述方式，深度揭示了软件定义网络的宏大家族及相关技术的嬗变史，从历史、发展、应用、未来、生态链、周边产业等多个角度剖析了软件定义网络。清晰准确地告诉你软件定义网络是什么，不是什么，它能帮你跳入软件定义网络这一浩瀚的海洋，同时又能避开迷思的漩涡。

本书宏观与微观兼具，树木与森林并览，读此书，你会惊讶于作者从如此多的视角来观察一项技术，给读者提供了充足的实例来认识一项技术的全貌。如果你是软件定义网络相关领域的用户或工程师，读此书也许可以解答困扰了你很长时间的疑惑，而针对这个疑惑你曾经从不同专家那里得到的却是完全相反的答案。读此书你会明白，这些答案只不过是从不同角度看问题的结果。

任何对网络和软件定义网络感兴趣的读者，都能无障碍地理解本书的内容，包括本科生和研究生、网络专业人员和 IT 经理、已有的和潜在的软件定义网络用户。本书内容包容性很强，不需要广泛的网络知识就能让读者充分了解其价值。本书通过对网络发展史的简介及其技术的描述，再加上对引发 OpenFlow 和软件定义网络行业全景的描绘，为读者提供了丰富的背景知识。如果你是一位网络架构师或 IT 经理，你经常会在市面上发现多种解决方案，它们的名字都是软件定义网络，但是其细节却大相径庭，阅读本书可以为你理解和评估市场上各种相互竞争的软件定义网络技术手段提供坚实的基础。

本书内容涵盖之广，超出你的想象，甚至包括软件定义网络对未来 IT 职业的影响！作为一个 IT 从业者，你的职业会受到软件定义网络的影响吗？你应该如何改变自己以适应这个发展大潮？建议你阅读本书。

限于水平，翻译不妥和错误之处在所难免。敬请广大读者批评指正。来信请发至 hai_wang@189.cn。

译 者

序　言

什么是软件定义网络？为什么它会成为网络界炙手可热的技术？对于一个诞生在五年前的术语来说，这个问题貌似简单，然而与“云计算”等其他一些术语一样，在一些营销团队为了达到自己的目的而对这些术语进行多年的重塑之后，已经变得越来越让人难以琢磨了。更加令人困扰的是新术语的出现，例如开放网络（open networking）、可编程网络（programmable network）和软件驱动网络（software-driven network）。由此引发很多人的好奇心，他们想知道 SDN 到底是什么？在其背后是否真的存在炒作？

对于像我们这样一直以来都在深度关注 SDN 发展的人来说，要想区分炒作与真相并不困难，我们也能够透过表象看到问题的本质及其来龙去脉。但是，从日常的授课情况来看，很多想要了解 SDN 的人正在被围绕这个话题产生的来自各个方面和各个角度的言论所困扰，这使得他们很难弄清楚真正发生的事情的核心所在。

2010 年 7 月，当我第一次公开谈论 SDN 的核心协议 OpenFlow 时，几乎还没有人听说过 OpenFlow 或软件定义网络。此后，我不断向数以千计的学生、网络工程师和以首席二字打头的高管们解释什么是 SDN。有些人在倾听了数分钟的解释之后立刻就“懂了”，此时往往我还没来得及放映第三张幻灯片，而另一些人则要花费整个全天课程的时间才能“明白”。比较普遍的现象是，能够最快理解 SDN 对行业所具有的潜在影响的人，正是那些网络界的元老，是那些能回忆起因特网前时代的人，那些实际上构建了这些网络产品的人，以及能够理解这些网络商业价值的人。

这就是为什么 Paul 和 Chuck 完全能够担当得起解释 SDN 所具有的价值的原因。他们都拥有非常丰富的网络从业经验，甚至可以追溯到比他们自己愿意承认的更为久远的时光。他们使用一代又一代的技术建立起各种网络解决方案。Paul 有两家成功的初创公司，他非常熟悉网络行业的各项业务工作。对于解释 SDN、定位 SDN 过去 30 年来在计算机网络界中的地位以及预测它在未来几年对网络具有的潜在影响来说，如此深厚宽广的从业经验绝对是无价之宝。

今天的因特网上并不缺乏有关 SDN 的信息，包括各种形式的博客、白皮书、播客和视频。然而，对于那些在过去几年里没有跟 SDN 打过交道而又想要加速赶上的人来说，我还没有发现有哪一本书能够像作者在本书中所做到的那样，从一个不偏不倚的角度来全面详尽地介绍 SDN。

任何读者只要对网络和 SDN 感兴趣，都可以轻松理解本书的内容，包括本科生和研究生、网络专业人员和 IT 经理。本书内容包容性很强，读者不需要丰富的网络知识就能充分认识到 SDN 的价值。它通过对网络发展史及其相关技术的简单介绍，以及对引发 OpenFlow 和 SDN 的行业全景的描绘，为读者提供了丰富的背景知识。如果你是一位网络架构师或 IT 经理，并且正

在试图比较多个都声称是基于 SDN 解决方案但又似乎是完全不同的技术时，本书可以帮助你理解和评估市场上互相竞争的各种 SDN 方案。

事实上，网络行业正在发生巨大的转变，而驱动这场变革的原因很简单，你可以从云计算及移动通信的兴起到 SDN（也可称之为开放网络或可编程网络，随便你怎么称呼）的发展之间直接画一条直线。如果你想绕过市场营销和炒作来对 SDN 做深入全面的了解，并弄清楚它是如何帮助重塑网络行业的，我绝对建议你阅读本书。

Matt Davy
Tallac Networks 公司
2016 年 4 月 26 日

Matt Davy 是软件定义网络技术的世界知名专家。在印第安纳大学，他曾担任 InCNTRE、SDN Interoperability Lab、Network Research 以及 Internships and Training 的执行主任。Davy 是拥有超过十二万用户、十万以太网端口和五千个无线接入点的企业网络的首席架构师，并且在大型网络运营商及企业网的设计和运行方面拥有 19 年的丰富经验。

前　言

当初，在我们着手构思本书时，部分动机是出于能够全面介绍软件定义网络（SDN）的读物少之又少。虽然作为作者，我们的专业就是与 SDN 打交道，但即便如此我们也不得不承认无法从某个单一的来源获得 SDN 的全面知识。我们意识到有相当数量的专业人员并没有直接接触 SDN，但他们很需要对其有所了解，这真是一个大问题。因此，简单地讲，本书的目的就是介绍产生 SDN 的环境，勾勒 SDN 区别于其他竞争技术的特点，并解释这种新兴技术已经展现出的众多重要的商业影响。为这样一种快速发展的技术撰写著作所面临的挑战正是这种技术不断发生的变化。

上面这段文字是为本书三年前出版的第一版所撰写的，不过今天看来仍然适用。很多没有在第一版中得到重视的技术现如今已被普遍承认是 SDN 的一部分。我们选择使用了“深度剖析”作为本书标题的一部分。有许多相互竞争的思想目前都在实际应用当中，而它们的开发者也都希望能够加入 SDN 的大潮。事实上，SDN 技术所涉及的范围也似乎在不断扩展。不论我们的读者需要应对哪些方面或哪种类型的 SDN 技术，至少我们希望通过阅读本书，读者能够在一个更加广泛的 SDN 环境之中对自己的需求加以考虑。为此，我们尝试讨论 SDN 的各种不同定义。本书在使用 SDN 定义时没有教条可寻，希望读者能够理解。

有兴趣了解软件定义网络或者对下面任何主题感兴趣的读者都会通过阅读本书有所收获：

- 组网
- 交换
- 软件定义网络
- OpenFlow
- OpenStack
- OpenDaylight
- 网络虚拟化
- 网络功能虚拟化

软件定义网络是一个正在迅速扩张的领域。虽然我们试图尽力考虑得全面且详尽，但感兴趣的读者或许还需要使用本书提供的参考资料对某些技术主题做更深入的探索。我们假设读者除对计算机的概念有基本了解外，并不具备专业知识。计算机编程和计算机组网的某些经验将会有助于理解本书的内容。本书包含了大量的数据和图表来解释和说明那些被定义或讨论的网络概念。这些图表帮助读者在不需要使用其他参考资料的前提下也能通读全书。

本书的第一版非常受欢迎。出版三年来，我们收到了许多以本书内容作为课程基础的大学教师的询问。第二版出版的动机包括以下两个方面：首先，本书在内容上需要紧跟 SDN 的发展而变化；其次，为了更加适用于 SDN 的研究生课程，我们做了一些针对性的调整。为此，我们在每一章都增添了一些文本框，其中包含的是与上文内容相关的讨论题。这些题目可以用来作为课堂讨论，也可以作为测验基础。此外，在链接 https://textbooks.elsevier.com/web/product_details.aspx?isbn=9780128045558 中可以注册下载与课程相关的辅导资料，如讲义和实验练习等。

建议和纠错

虽然我们已经尽可能细心，但本书仍然会有一些错误，并且某些读者感到尤为关心的主题也有可能被忽略了。我们期待在未来的版本中能够纠正所有的错误，并包容尽可能多的建议。如果有什么建议，请通过以下电子邮件发送给我们：chuck.a.black@gmail.com。

作者简介

Paul Göransson 是一位有多次创业经历的企业家，他的两家初创公司都已被业界巨头收购：惠普收购了 Qosnetics（1999 年）；思科收购了 Meetinghouse（2006 年）。Göransson 曾在安捷伦科技（Agilent Technology）公司的先进网络部门和思科（Cisco）公司的无线网络业务部担任高管，现任 Elbrys Networks 公司的创始人和主席。Göransson 在 1975 年获得美国布兰迪斯大学心理学学士学位，并分别于 1981 年和 1995 年获得美国波士顿大学计算机工程硕士学位和美国新罕布什尔大学计算机科学博士学位。Göransson 是一名狂热的马拉松运动员、登山家、铁人三项选手，并且一直是活跃的潜水和户外运动爱好者。他参加过上百次马拉松比赛、数次铁人三项运动赛以及多次超级马拉松比赛，包括 Leadville Trail 100，同时他还是一名 PADI 潜水教练。2015 年，Göransson 依靠他新装的人造右脚踝完成了长达 2189 英里的阿帕拉契小径（Appalachian Trail）的徒步旅行。Göransson 曾在法国、阿尔及利亚、委内瑞拉、墨西哥和瑞典居住、学习和工作了很长时间。Göransson 与人合著了 *Roaming Securely in 802.11 Networks*，即《802.11 网络安全》，并在计算机网络相关的多家期刊发表多篇文章。他也经常应邀在技术会议上发表演说。在业余时间里，Göransson 拥有并管理着美国缅因州南部的一个 220 英亩的无角海福特牛（Polled Hereford）与牧草农场。

Chuck Black 在计算机网络领域拥有超过 35 年的工作经验，其中大部分时间都在惠普公司的研究与开发实验室工作，之后成为 Tallac Networks 公司的联合创始人之一。Tallac Networks 公司是一家以软件定义网络为产品的初创公司。Black 最近在 SDN Essentials 公司对其主要网络供应商旗下的工程师和客户进行 SDN 应用程序开发领域方面的培训。Black 是惠普公司网络接入控制和安全领域多种网络产品的创新者和开发者，并且在这些领域拥有 11 项专利。而在此之前，他曾为惠普软件部门开发过一些网络管理方面的产品。在局域网发展的初期，他是业界最早出现的网络拓扑发现协议的开发者之一。Black 拥有美国加利福尼亚理工大学计算机科学学士学位和硕士学位。

Timothy Culver 在信息工程、信息技术、业务拓展和市场销售等各个方面都拥有大量工作经验，是一位经验丰富的技术主管和大学教师。他对新兴技术充满热情，并且是一位成功的实践者。他在组建全球技术团队方面拥有良好的业绩。Culver 曾经经历了 3 个初创企业的创建工作，并建立了跨 14 个国家的业务。Culver 作为美国得克萨斯大学达拉斯分校计算机科学与软件工程研究生课程的教授，率先推出了有关软件定义网络的课程及项目。在 20 世纪 90 年代任教于美国得克萨斯大学达拉斯分校之前，Culver 曾在达拉斯社区学院教授计算机语言和计算机科学入门课程。Culver 因积极参与社区志愿者活动而获得 2011 年至 2014 年的美国总统志愿者服务奖，这些活动包括针对初中和高中学生的 STEM（Science Technology Engineering and Math）教学。他积极帮助学生们迈入相关领域的职业生涯，并且还为 WeTeachScience.org 的导师咨询委员会提供服务。在个人生活方面，超过三十年的婚姻生活使他成为四个孩子的父亲。作为孩子们的教练，Culver 在踢了 25 年足球后取得美国足球联合会国家教练和美国青年足球国家教练执照。Culver 还在美国和欧洲参加过多场足球比赛。目前 Culver 居住在得克萨斯州桑尼维尔的一个农场里，在那里他饲养着一群安格斯牛。Culver 在 VoIP、互联网电话会议和 LDAP 领域拥有 4 项专利。他是美国贝勒大学的荣誉毕业生，并于此获得 MIS 学位。之后，Culver 在美国南卫理公会大学获得工程硕士和工商管理硕士学位。他在美国贝勒大学和美国瓦尔登大学进行云计算专业的研究生学习。Culver 还是开放网络基金会的研究助理。

致　谢

非常感谢这些年来我们的家庭给予我们的大力支持，特别是在撰写本书期间给予我们的支持。

如果没有 Bill Johnson、Matt Davy 和 Paul Congdon 博士坚持不懈的支持，就不可能有本书的问世，他们都是 Tallac Networks 公司的联合创始人。他们对 SDN 技术的深度理解以及克服各种困难对本书手稿所做的审阅工作，还有对本书的其他许多直接贡献都是不可估量的。

我们要感谢来自荷兰 Sicse 的互联网通信与服务工程师 Niels Sluijs 博士，他对本书的手稿给出了许多令人耳目一新并且非常中肯的观点和评论。

我们还要感谢远在泰国清迈大学的 Nopadon Juneam，他帮助我们为书中多幅图片的最终版本定型。

我们还要感谢 Tallac Networks 公司的 Ali Ezzet，他仔细阅读了手稿的各个章节，并运用自己的技术专长为我们检查出不少错误，大大提高了稿件的质量。我们同样感谢 Anthony Delli Colli 在商业和市场营销方面为本书提供的许多重要信息和建议。对本书来说这些都是不可或缺的贡献。

特别感谢 Helen Göransson 为本书手稿所做的多次精心审阅。她的努力显然使得本书更具可读性。

除以上各位对本书第一版所做的宝贵贡献外，我们还要感谢对本书第二版进行仔细审阅的两位专家，他们使得第二版有了相当大的改进。为此，衷心感谢思科公司的 Giles Heron 以及来自 Rio Grande do Sul 联邦大学（UFRGS）的 Alberto Schaeffer-Filho 博士。

最后，特别感谢 Elsevier 的 Brian Romer 和 Amy Invernizzi 对这个项目的鼓励和支持。

我们也非常感谢 Tallac Networks 公司为全书使用插图的选定所做出的贡献。

Paul Göransson
Chuck Black
Timothy Culver

目　录

Chapter 1

引　言

作为一名科技文章的作者，很少有机会能够在新近出版的头条新闻中看到自己所要撰写的主题，围绕软件定义网络（Software Defined Networking，SDN）刮起的强劲风暴却使得它冲上了头条[1]。当代的计算机网络已经演变成一个对管理极具挑战的可怕怪兽，并且仍然在不断膨胀以期满足当前环境下的诸多要求。SDN 代表了一种试图解决目前网络范式存在缺陷的新手段。SDN 是给交换机编程以将其应用于现代数据网络的一种创新方式。SDN 的发展方向是高度可扩展的集中式网络控制体系结构，它非常适用于当前数据中心普遍存在的超大规模网络。SDN 的设计初衷是执行细粒度的流量转发判决，而不是把应用程序特定的转发行为强塞给并不适合这项任务的传统体系结构。目前，人们对这种新兴的互联网交换技术的兴趣已经远远超出了学术研究和技术社区的范围。近三十几年来，互联网技术的发展仅仅体现在一些断断续续的修修补补上，这使得互联网技术陷入了发展僵局（有时也称为互联网骨化），这也是 SDN 的早期驱动因素之一，它通过促进实验及开发新协议和应用程序来解决这一问题。如果 SDN 的技术承诺得以实现，那么它所代表的不仅是网络行业的结构性转变，老牌的行业巨头也很有可能会被拉下台，而被转嫁给消费者的成本也有可能会暴降。不过，人们的期待也确实存在某种程度上的夸张，我们不仅需要了解这种新的网络范式所具有的潜能，也应当了解其局限性，这一点很重要。本书尽量从技术的角度解释 SDN 是如何工作的，并综述哪些网络应用是它所适用的，而哪些并不适用，我们还要讲解如何在此技术之上构建定制的应用程序，并从网络业务本身来探讨 SDN 所具有的影响力。

作为引言，本章将介绍当前互联网交换机的基本概念及其背景知识，包括对数据平面、控制平面和管理平面的定义和讨论。这些概念是理解 SDN 如何通过与传统交换机体系结构截然不同的方式实现其核心功能的关键。本章还解释了目前的网络实现是怎样执行转发判决的，并说明这种缺乏灵活性的转发判决使得网络管理员很难根据不同的条件对网络进行优化。本章继而通过举例说明为什么更加灵活的转发判决可以大大提高现有交换机的业务多样性，并解释为什么将控制平面从交换机中分离出来，使其成为一个具有独立、开放平台的控制器后，就能提高转发判决的灵活性。通过将其与 Linux 操作系统进行类比而后得出的结论

http://dx.doi.org/10.1016/B978-0-12-804555-8.00001-6

是：Linux 操作系统因充分利用了开源社区得到了迅速发展，如果将同样的模式应用到互联网交换机控制平面的开发上，也将具有相同的效果。

下面，我们先介绍一些与分组交换技术相关的贯穿全书的基本术语，然后再概述分组交换技术及其发展历程。

1.1 分组交换的基本术语

本节定义了与分组交换技术相关的一些基本术语，这些术语的使用将贯穿全书。我们的惯例是使用楷体字来强调新术语的首次使用。对于本节未定义的那些专业性更强的术语，将会在它们被首次使用时再行定义。许多分组交换术语和短语对于不同群体而言具有不同的含义，本书尝试对这些术语和短语使用人们最为普遍接受的定义。首字母缩写的简称也会在首次使用时定义并强调。在本书末，按字母顺序列出了书中用到的所有首字母缩略语。熟悉网络的读者可以跳过本节，而其他一些读者可能希望先通读本书的全部内容，然后再回过头来参考特定的术语。

在解释 SDN 与传统分组交换网络有何不同时，术语是一个很重要的参考，虽然 SDN 在某种程度上已经背离了这些传统的概念，或者从根本上改变了这些概念的含义。在阅读本书的过程中，我们鼓励读者随时回顾这些术语的定义，并思考这些术语的含义在 SDN 中何时保持未变，SDN 何时需要更细致的定义，以及在 SDN 的讨论中何时需要用到全新的词汇。

广域网（Wide Area Network，WAN）指覆盖广阔地理区域范围的网络，通常包含不止一个城市区域。

局域网（Local Area Network，LAN）指覆盖有限地理区域范围的网络，面积一般不会超过几千平方米。

城域网（Metropolitan Area Network，MAN）是填补局域网和广域网之间空隙的网络。人们之所以开始使用这个术语，是因为局域网和广域网最初的区别不仅仅体现在地理覆盖范围上，它们所使用的传输技术以及速率也不同。但是随着既有类似局域网的传输速率和接入控制机制，又能够服务于城市大部分区域的技术的出现，人们开始使用术语“城域网”来表示这样一种既不同于大型局域网，也不同于小型广域网的新实体。

无线局域网（Wireless Local Area Network，WLAN）是以空气为传输媒体的局域网。在无线网中，任意两个设备之间的最大距离通常在 50 米左右。虽然无线通信也可以不使用空气作为传输媒体，但在本书中我们不考虑。

物理层（Physical Layer）是七层计算机网络*开放系统互连*（Open System Interconnection，OSI）模型[2]的最底层。它由基本的硬件传输技术组成，被用来在网络中传输数据比特。

数据链路层（Data Link Layer）是 OSI 模型的倒数第二层。这一层所提供的是在同一个网段内将数据从一个设备传送到另一个设备的能力。为了清楚起见，我们把一个 LAN 网段等同于一个冲突域。LAN 网段的严格定义是指网络设备之间的电气或光学连接。在我们对数据链路层的定义中，通过中继器连接的多个网段也被当作是一个 LAN 网段。网段的例子包括像以太网这样的单个 LAN，或者是广域网中相邻结点之间的点对点的通信链路。数据链路

层包括：（1）用于检测传输过程中可能出现的序列差错或比特差错的机制；（2）用于在同一个网段相连的发送器和接收器之间控制流量的机制；（3）允许多种网络协议使用同一通信媒体的复用能力。以上三个功能被认为是属于数据链路层的逻辑链路控制（Logical Link Control，LLC）部分，而数据链路层的其余功能则属于下面将要单独介绍的媒体接入控制（Media Access Control，MAC）部分。

媒体接入控制属于数据链路层的一部分，它在接入共享媒体时实施控制，并在多个接收器都能接收到数据但只能由其中一个接收器进行处理的情况下提供寻址功能。就本书而言，我们不区分数据链路层和 MAC 层。

网络层（network layer）提供了一些功能和过程，以允许数据从发送方跨越多个中间网络传输到达接收方。每个中间网络的传输都会涉及上文所述的数据链路层的处理过程。网络层则负责将这些分散的过程黏合到一起，使数据能够从发送方正确抵达预期的接收方。

一层（layer one）的定义与上述物理层的定义相同。

二层（layer two）的定义与上述数据链路层的定义相同。我们也会使用术语 L2，它与二层同义。

三层（layer three）的定义与上述网络层的定义相同。在本书中，L3 与三层同义，并可互换使用。

端口（port）是指与某个通信媒体之间的连接，它包括一整套的数据链路层和物理层的机制，这些机制是通过该链路正确传送和接收数据所必需的。链路可以是任何可能的媒体类型。在本书中，端口可以与同义术语接口（interface）互换使用。因为本书还涉及虚拟交换机，所以端口的定义也可以扩展为虚拟接口（virtual interface），也就是隧道的端点。

帧（frame）是指在二层网络上传输的数据单元。

分组（packet）是指在三层网络上传输的数据单元。通常，在不刻意区分二层还是三层的时候，这个术语也用于指代在二层网络上传输的数据单元（帧）。当二层和三层之间的区分很重要时，分组通常是帧的有效载荷。

MAC 地址（MAC address）是能够在全球范围内唯一标识某个网络设备的一个数值。虽然 MAC 地址是全球唯一的，但它们只能用于二层地址，所标识的是二层网络拓扑结构上的一个设备。

IP 地址（IP address）是给计算机网络中使用了 IP 协议的每台主机分配的一个名义上的唯一值，用于三层寻址。

IPv4 地址（IPv4 address）是符合 IPv4 协议规则的 32 位整数值的 IP 地址。这个 32 位整数值通常以点分（dotted）记数法表示，其中在组成地址的 4 个字节中，每个字节都用 1 个 0 到 255 之间的十进制数来表示，并以句点分隔（如 192.168.1.2）。

IPv6 地址（IPv6 address）是符合 IPv6 协议的 128 位整数值的 IP 地址，其地址空间要比 IPv4 大得多。

交换机（switch）是一种多端口设备。交换机从一个端口接收信息，并将信息从另一个或多个端口发送出去，从而使这些信息传输到达指定目的地。

电路交换机（circuit switch）是这样的一种交换机，所有指定如何转发某条电路（即连接）上的数据的相关信息都会在交换机中保留一段规定的时间，即使该连接上有时候并没有需要处理的数据。这些相关信息的建立，要么是通过配置实现的，要么由每种电路交换机所指定

的呼叫建立（call set-up）或连接建立（connection set-up）的过程来实现。

分组交换机（packet switch）是这样一种交换机，构成两个或多个实体之间通信的数据被当作一个个独立的分组进行处理，每个分组自行穿越网络并到达目的地。分组交换机可划分为面向连接的和无连接的两种类型。

在面向连接的（connection-oriented）模式下，当数据途经网络时，每个中间交换机都驻留了一些相关的上下文信息，使得该交换机能够将数据转发到达其目的地。前文所述的电路交换机就是一个面向连接模式的很好的例子。

在无连接的（connectionless）模式下，数据通过网络时，每个分组中都包含了足够的信息，使得中间交换机能够在没有任何先验的上下文信息的情况下把数据转发到达目的地。

路由器（router）是子网和子网之间的分组交换机。子网（subnet）是由一组共享相同网络前缀的主机组成的网络。网络前缀由 IP 地址的前几位构成，且前缀的长度可变。通常，一个子网的所有主机位于同一局域网内。现在，术语“路由器”经常与“三层交换机（layer three switch）”互换使用。家庭无线接入点往往将 WiFi、二层交换机以及路由器的功能集成到一个机盒里。

洪泛（flood）是指交换机将分组从除接收该分组的那个端口外的所有端口发送出去。

广播（broadcast）与分组的洪泛相同。

线速度（line rate）是指连接到交换机端口的通信媒体的带宽。现代交换机的带宽通常以兆比特每秒（Mbps）或吉比特每秒（Gbps）为度量单位。当我们说一个交换机以线速度处理分组时，就意味着它能够处理以该带宽连续到达的分组流。

WiFi 是基于 IEEE 802.11 标准的无线通信系统的通用名称。

1.2 历史背景

20 世纪上半叶，世界各国的主要通信网络都是电话网。这些电话网普遍为电路交换网络。端点与端点之间的通信过程包括在通话之前先建立一条通路，然后在通话结束后拆除该通路。而在通话期间，话音信号所经过的通路是静态的。这种类型的通信也称为面向连接的通信。除了基于电路交换这个特点，这些电话网在很大程度上都是集中式的，也就是大量的终端用户连接到一个大型的交换中心。Paul Baran——一位在 20 世纪 60 年代成为美国 Rand 公司研究人员的波兰移民——认为一旦敌人发动攻击，像电话网这样的网络很容易受到破坏[3, 4]。网络生存能力较差的特点表现为某一个交换中心的损毁就有可能导致这个国家的大片区域失去电话通话功能。Baran 提出了以分组形式传输话音信号的解决方案，这些分组可以自主地穿越网络到达目的地。此概念所包含的思想是，如果通话路径的一部分遭受敌人的攻击，那么使用其他的替代路径自动重新选择路由也能到达相同的目的地，从而使这个通话过程仍然保持。Baran 证明即使有 50%的转发交换机被损毁，全美的话音通信系统仍然可以正常工作，从而大大降低了当时非常普遍的集中式电路交换体系结构所具有的脆弱性的特点。

当 Baran 在 Rand 公司埋头苦干时，他可能从来也没想到过自己的核心思想最终会在数

据网络领域得以实现。尽管这在当时肯定不是一种被普遍接受的网络模式，但其后的历史已成为当前互联网传奇的一部分。Baran 的想法在美国国防部（DOD）于 1969 年开始进入实际运行的实验性 ARPANET 网络中得以体现。ARPANET 连接了学术研究机构、军事部门和国防承包商。这种分布式的无连接网络模式经过多年发展，最终在 20 世纪 90 年代以因特网的形式异军突起，成为当今世界上最著名且最受欢迎的商用网络。

在 ARPANET 出现后的数十年里，网络界专业人士就基于连接的还是无连接的、集中式的还是分布式的体系结构哪一种更有优势进行了长期不懈的拉锯战。在双方争执的过程中，有时候形势似乎倾向于某一方，但没过几年，形势又被逆转。20 世纪 90 年代，因特网的爆炸式成长似乎为此争论画上了一个休止符，至少就计算机网络而言是如此。不断崛起的因特网显然是一个分布式的、无连接的体系结构。类似 X.25[5]这种面向连接的协议似乎注定要成为过去式。任何过于集中化的设计都被认为是非常易于受到攻击的，不论是恶意攻击，还是一次简单的自然现象导致的后果。即使像异步传输模式（Asynchronous Transfer Mode，ATM）[5]这样的新兴产品，虽然在 20 世纪 90 年代中期被鼓吹为能够处理的线速度大大超过了因特网的能力范围，但最终还是大部分被比以往更加灵活的因特网所取代。因特网在做出一些改变之后居然也可以处理每秒几十吉比特范围的线速度，而这曾经被认为只有通过像 ATM 之类的信元交换技术才有可能实现。

讨论题：

请说明为什么 Paul Baran 对无连接网络的看法会吸引美国国防部的关注。

1.3 现代数据中心

因特网对计算机网络的统治程度在科技史上是罕见的，而其核心思想就是无连接的和分布式模式，而在因特网的发展过程中又衍生出了万维网（World Wide Web，WWW，由英国 Tim Berners-Lee 爵士始创）。由于数据中心托管的 Web 订阅服务越来越复杂、越来越繁重，这使得数据中心的规模不断膨胀。数据中心是一个从物理上保护起来的大型机房，内部存放了大量用于计算和存储的计算机。通过选址在不太可能出现自然灾害的地理位置上，并使用冗余的电源系统，以及通过在不同地点进行热备份等，这些大型机房使得自身尽可能地受到保护，以最大限度地免受环境影响。

由于服务器数量庞大，因此它们被安放在高度有组织且成排放置的服务器机架上。经过一段时间的发展，出于效率上的考虑，独立的服务器逐渐演变成密集架上的刀片（blade）服务器。服务器机架的分级组织模式使得顶架（Top-of-Rack，ToR）交换机需要提供机架内部的组网以及机架之间的接口功能。图 1.1 描绘了目前数据中心普遍采用的严格的分级结构。

在发展过程中，仅从计算机的数量上就能看出数据中心扩张的迅猛势头。目前正在建设的数据中心可以容纳超过 120 000 台物理服务器[6]。对于最新的物理服务器来说，每台服务