计算机科学丛书

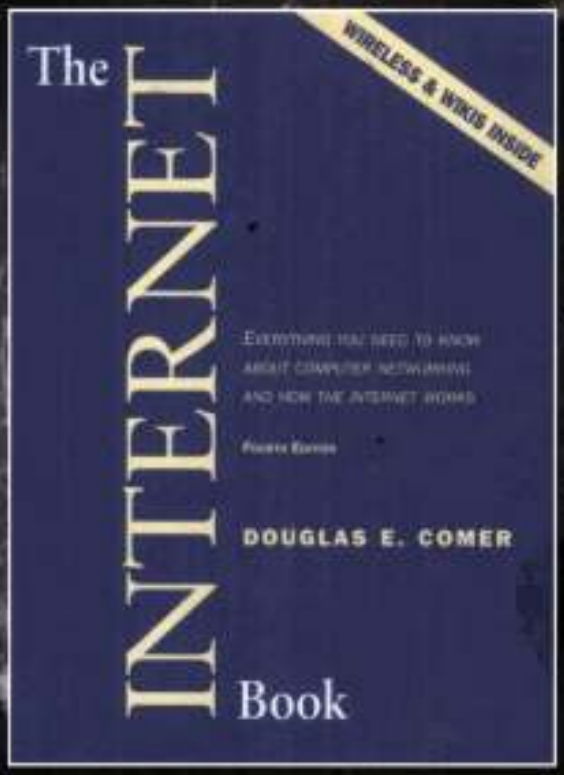
原书第4版

Internet技术基础

(美) Douglas E. Comer 著
Cisco公司&普度大学

兰小丰 李潇 陈志 译
南开大学

吴功宜 吴英 审校
南开大学



The Internet Book

Everything You Need to Know about Computer Networking and How the Internet Works

Fourth Edition

机械工业出版社
China Machine Press

Internet技术基础（原书第4版）

"在可读性和清晰度方面，Douglas的表达力无出其右。本书的写作风格直接而且重点突出。本书是提供给网络初学者的一本优秀教材。"

——Raymond Hsieh，纽约州立大学布法罗分校

"本书以一种简洁的语言向学生介绍了因特网，指出了关键所在并预见未来。"

——RobinPeek，西蒙斯学院

Douglas Comer是著名的网络技术作家、教师和顾问，他对因特网的概念和工作原理有独到的见解和清晰的认识。在本书第4版中，Comer教授从一种非技术化的角度来解释因特网后台的工作机制、计算机的通信方式、服务器以及因特网技术的原理。作者借助实用的方法，以清晰易懂且科学准确的语言介绍了因特网的各种技术及其不同的应用，向读者展示出网络互连和因特网的全景图。

本书主要特色：

- 三大新增章节：
  - NAT：共享Internet连接
  - 远程安全接入（VPN）
  - 群组和个人网页（维基及博客）
- 关键网络概念，如分组交换、局域网、协议软件和域名。
- 全新的内容，包括Wi-Fi的无线联网技术、Web制作工具，以及关于网络安全性的讨论。
- 包括电子商务的概述以及如何利用网络进行业务和商业交易。
- 采用更多易于理解的定义扩展了术语表的内容。

作者简介

Douglas E. Comer Cisco公司研发副总裁，同时还是普度大学计算机科学系教授。他曾是因特网体系结构委员会的成员，该委员会是确定因特网发展标准的权威机构；同时还是美国计算机学会的会员。除本书外，他还著有《TCP/IP网际互联》（3卷本）、《计算机网络和因特网》等。

PEARSON HALL www.PearsonEd.com

投稿热线：(010) 88379604
购书热线：(010) 68995259, 68995264
读者信箱：hzjsj@hzbook.com

华章网站 http://www.hzbook.com

网上购书：www.china-pub.com

上架指导：计算机/网络通信

ISBN 978-7-111-23789-1

9 787111 237891

定价：30.00元

计 算 机 科 学 丛 书

Internet技术基础

(美) Douglas E. Comer 著 兰小平 李涛 陈志 译 吴功宜 吴英 审校

The Internet Book

Everything You Need to Know about Computer Networking and How the Internet Works

Fourth Edition

机械工业出版社
China Machine Press

本书阐释了计算机之间的通信原理、Internet 的本质、Internet 的运行方式以及 Internet 能够提供哪些服务。主要内容包括：电话技术、模拟技术、数字网络、基本通信、局域网、Internet 的发展、包交换技术、宽带与无线接入技术、IP、TCP、分布式计算、NAT、电子邮件、电子公告板服务、HTML、高级网页技术、搜索引擎、IM 和 VoIP、FTP、TELNET、VPN、电子商务和数字化图书馆等内容。

本书内容详尽，适合高校本科生作为网络教材，也适合网络初学者作为网络入门参考书使用。

Simplified Chinese edition copyright ©2008 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *The Internet Book, Everything You Need to Know About Computer Networking and How the Internet Works Fourth Edition* (ISBN 0-13-233553-0) by Douglas E. Comer, Copyright ©2007.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

本书封面贴有 Pearson Education（培生教育出版集团）激光防伪标签，无标签者不得销售。

版权所有，侵权必究。

本书法律顾问　北京市展达律师事务所

本书版权登记号：图字：01-2006-6510

图书在版编目（CIP）数据

Internet 技术基础（原书第 4 版）/（美）科默（Comer, D. E.）著；兰小丰等译. —北京：机械工业出版社，2008.4

（计算机科学丛书）

书名原文：The Internet Book, Everything You Need to Know about Computer Networking and How the Internet Works, Fourth Edition

ISBN 978-7-111-23789-1

Ⅰ. I…　Ⅱ. ①科…　②兰…　Ⅲ. 因特网　Ⅳ. TP393.4

中国版本图书馆 CIP 数据核字（2008）第 040762 号

机械工业出版社（北京市西城区百万庄大街 22 号　邮政编码　100037）

责任编辑：王　璐

山西新华印业有限公司新华印刷分公司印刷・新华书店北京发行所发行

2008 年 5 月第 1 版第 1 次印刷

184mm×260mm・14 印张

标准书号：ISBN 978-7-111-23789-1

定价：30.00 元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换

本社购书热线：（010）68326294

出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅奠定了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切，这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短、从业人员较少的现状下，美国等发达国家在其计算机科学发展的几十年间积淀的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章图文信息有限公司较早意识到"出版要为教育服务"。自1998年开始，华章公司就将工作重点放在了翻译，移译国外优秀教材上。经过几年的不懈努力，我们与Prentice Hall，Addison-Wesley，McGraw-Hill，Morgan Kaufmann等世界著名出版公司建立了良好的合作关系，从它们现有的数百种教材中甄选出Tanenbaum，Stroustrup，Kernighan，Jim Gray等大师名家的一批经典作品，以"计算机科学丛书"为总称出版，供读者学习、研究及收藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

"计算机科学丛书"的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，"计算机科学丛书"已经出版了近260个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍，为进一步推广与发展打下了坚实的基础。

随着学科建设的初步完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都步入一个新的阶段。为此，华章公司将加大引进教材的力度，除"计算机科学丛书"之外，对影印版的教材，则单独开辟出"经典原版书库"。为了保证这两套丛书的权威性，同时也为了更好地为学校和老师们服务，华章公司聘请了中国科学院、北京大学、清华大学、国防科技大学、复旦大学、上海交通大学、南京大学、浙江大学、中国科技大学、哈尔滨工业大学、西安交通大学、中国人民大学、北京航空航天大学、北京邮电大学、中山大学、解放军理工大学、郑州大学、湖北工学院、中国国家信息安全测评认证中心等国内重点大学和科研机构在计算机的各个领域的著名学者组成"专家指导委员会"，为我们提供选题意见和出版监督。

这两套丛书是响应教育部提出的使用外版教材的号召，为国内高校的计算机及相关专业的教学度身订造的。其中许多教材均已为M. I. T.，Stanford，U.C. Berkeley，C. M. U. 等世界

名牌大学所采用。不仅涵盖了程序设计、数据结构、操作系统、计算机体系结构、数据库、编译原理、软件工程、图形学、通信与网络、离散数学等国内大学计算机专业普遍开设的核心课程，而且各具特色——有的出自语言设计者之手，有的历经三十年而不衰，有的已被全世界的几百所高校采用。在这些博学渊博的名师大作的指引之下，读者必将在计算机科学的宫殿中由登堂而入室。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。但我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。教材的出版只是我们的后续服务的起点。华章公司欢迎老师和读者对我们的工作提出建议或给予指正。我们的联系方法如下：

电子邮件：hzjsj@hzbook.com
联系电话：(010) 68995264
联系地址：北京市西城区百万庄南街1号
邮政编码：100037

专家指导委员会

（按姓氏笔画顺序）

尤晋元	王　珊	冯博琴	史忠植	史美林
石教英	吕　建	孙玉芳	吴世忠	吴时霖
张立昂	李伟琴	李师贤	李建中	杨冬青
邵维忠	陆丽娜	陆鑫达	陈向群	周伯生
周克定	周傲英	孟小峰	岳丽华	范　明
郑国梁	施伯乐	钟玉琢	唐世渭	袁崇义
高传善	梅　宏	程　旭	程时端	谢希仁
裘宗燕	戴　葵			

译 者 序

21世纪的一个重要特征是数字化、网络化与信息化，而它的基础是支持全世界的强大的计算机网络。计算机网络技术已经对社会发展和科学技术的进步产生了不可估量的影响。以Internet为代表的网络应用技术已经成为计算机应用中一个空前活跃的领域。计算机网络已经和电力、电信一样，成为支持现代社会整体运行的基础设施，成为人们须臾不能离开的东西。基于Web技术的电子政务、电子商务、远程教育、远程医疗正在以前所未有的速度发展，搜索引擎、博客、播客、即时通信、网络电话、网络电视、网络游戏等新的应用正在迅速普及。计算机网络正在改变着人们的工作方式与生活方式，网络技术的发展与应用已成为影响一个国家或地区政治、经济、科学与文化发展的重要因素之一。

Internet的发展速度是非常惊人的，并且不断有新的服务类型涌现出来。根据2007年6月中国互联网络信息中心（CNNIC）发布的第20次《中国互联网络发展状况统计报告》数据显示：我国网民的人数已经达到1.62亿，仅次于美国的2.11亿。但是从Internet普及率的角度，我国的Internet普及率只有12.3%，低于全球平均的Internet普及率16.3%，与世界发达国家或地区相比还存在着一定的差距。从我国经济与社会发展的大好形势看，未来的中国Internet普及率将会呈现大幅度提升的局面。因此，普及Internet知识是摆在全社会面前的一项重要任务。

本书作者Douglas E. Comer是国际著名的网络技术专家，他的很多著作已经被翻译成多种文字，在世界各国发行。他对Internet的概念和工作原理有独到的见解。在本书中作者用清晰的层次结构、准确的概念与通俗的语言，为读者生动和准确地剖析了Internet的工作机制与应用。对于计算机网络与Internet背景知识不足很多的读者，或者是讲授计算机网络和Internet课程的教师来说，这是一本非常适宜的入门书籍和参考读物。通过这本书，读者可以循序渐进地学习Internet基础知识以及了解Internet技术的一些新的发展动态。

本书的序言与第1章～第14章由兰小丰翻译，第15章～第24章由李鑫翻译，第25章～第34章由陈志翻译，全书由吴功宜、吴英审校。我们在翻译本书的过程中希望能尽可能地尊重原著的思想，但是限于译者的学识，书中难免存在疏漏与理解错误之处，敬请读者指正。

译 者

2007年年底

于南开大学信息技术科学学院

网络实验室

前言

本书阐释了计算机之间是如何通信的，什么是Internet、Internet如何运行以及Internet能够提供哪些服务。本书针对的是没有较强技术背景的读者，书中前面的一些章节为读者清晰地解释了有关服务的术语和概念，这些都是读者需要提前了解的内容。读完本书之后，你将理解Internet所包含的各种技术，也会惊叹Internet在各方面的应用。你将发现Internet为什么能够让人们如此兴奋。除此之外，你还将明白Internet的起源以及它是如何迅猛发展的。

本书通过源于日常生活的例子，用类比的方法来解释技术，而并非采用数学方法、算法或者计算机程序来达到这个目的。例如，为了解释为什么数字通信优于模拟通信，书中给出了在雾气中用闪光灯发送信号的类比；为了说明声音如何能够以一种平稳的速度播放给用户，书中用到了这样一个类比：超市每次进货总是一次买入许多加仑的牛奶，卖出时却是一次一加仑。

除了讲解用户用到的诸如电子邮件、文件下载、即时通信和Web浏览等服务，本书还包括了所有至关重要的网络概念，比如包交换技术、局域网技术、协议软件和域名技术。更重要的是，本书贯彻了这样一个基本原则——先描述基本的Internet通信设施，然后介绍这些基础设施是如何用于提供各种各样的服务的。另外，书后还列有一张易于读者理解的术语表，鼓励读者在阅读本书的时候参考。

本书第4版保留了第3版的总体框架结构，在此基础上增加了3个新章节（分别是第19、26和32章），更新了若干素材资料。第19章阐述网络地址转换（NAT）技术，许多Internet用户在家里使用这项技术；第26章说明了两个新的Internet应用——博客和维基；第32章，第3个新增章节，解释了虚拟专网（VPN）技术，该技术保证雇员可以从远程任何位置安全地访问公司网络。

与上一版一样，本书第4版被分成四部分。第一部分讲的是一些基础概念，如数字和模拟通信。除此之外，该部分还介绍了包交换技术，解释了被广泛应用于商业领域的局域网技术。

第二部分叙述了Internet研究工程以及Internet发展的简史。尽管这段历史的大部分会被略过，但是读者应该注意到其超常规发展的增长率，该增长率表明这项技术在设计上取得的令人难以置信的成功，没有其他那项通信技术能在如此迅猛的发展过程中一直保持不变。

第三部分解释了Internet如何运行，其中描述了两个被所有服务用到的基础协议：网际协议（IP）和传输控制协议（TCP）。虽然该部分章节忽略了技术细节，但是它们可以帮助学生明白每个协议的基本作用，了解协议整体的设计思想。

第四部分剖析了可用的Internet服务。除了讨论万维网上使用的浏览器、Web文档和搜索引擎之外，这一部分还讨论了电子邮件、电子公告板、文件传输、远程桌面、维基、博客以及音频和视频通信。通过多个案例解释了Internet服务是如何运作的，以及它如何使用下层的设备。第四部分的结尾讨论了网络安全、虚拟专用网和电子商务技术。

本书可以作为高校网络课程极好的参考教材。虽然本书是以非技术的方式呈现给读者的，但是书中各方面的资料、信息是正确的和科学的。更为重要的是，在21世纪，求知者不仅需要知道如何使用浏览器或者如何构建Web页面，他们还应该明白这些表象下面发生了什么。他们可以从本书获取这些知识。

笔者鼓励教师把课堂讲授和实验教学结合起来，因为实验可以让学生直接看到和亲手使用这些技术。在整个课程教学中，先期实验应该把探索各种服务作为重点，其中包括发送电子邮件、使用浏览器和搜索引擎、下载文件、听取音频信息以及使用 IP 电话。当然这些实验内容的前提是具有相关实验环境。同时，笔者也鼓励所有学生，甚至包括那些对计算机技术不感兴趣的学生，亲手制作一个简单的 Web 页面。网页制作除了可以帮助他们明白 HTML 文档中的标签与结果页面显示之间的关系，还可以向学生展示服务器是如何把文件从本地硬盘传送到浏览器上的。了解各个实验之间的关系有助于使学生更好地明白他们从书本中了解到的底层机制。

后期实验项目依课程类型而定。面向商业的课程往往重点要求学生使用 Internet 或者从事案例研究。实验需要学生在 Internet 上查找一些信息，然后写一篇论文对这些信息进行分析。一些其他的课程通过实验学习使用诸如制作网页的工具。还有些课程把学生查找信息和制作包含链接到这些信息的网页这两方面的内容结合起来。无论如何，我们已经发现学生是带着真正的热情和动力融入到 Internet 课程中的，教授的任务仅仅是提供一些观点和见解，通过本学期的教学让学生明白，为什么 Internet 如此令人着迷。

笔者感谢对这本书的每个版本做出贡献的人们。感谢 John Liu、Keith Rovell、Rob Slade 和 Christopa Schube 阅读了该书的早期版本并提出了建议。感谢 Dwight Barnette、George Polyzos、Donald Knudson、Dale Musser 和 Dennis Ray 为前一版撰写评论。感谢 Scott Comer 提出了一些看法。最后，要感谢我的夫人 Christine 仔细编辑了本书的手稿，解决了许多问题，并润色了文中的措辞。

Douglas E. Comer

2006 年 5 月

目录

出版者的话
专家指导委员会
译者序
前言

第1章 Internet时代已经到来 …… 1
1.1 世界已经改变 …… 1
1.2 数字不能说明问题 …… 2
1.3 学习Internet …… 2
1.4 了解Internet的概念 …… 2
1.5 术语和技术 …… 2
1.6 Internet的发展及其适应性 …… 3
1.7 Internet的影响 …… 3
1.8 本书的组织结构 …… 3
1.9 个人见解 …… 3
第2章 开始吧：亲身体验 …… 4
2.1 引言 …… 4
2.2 Web：网站与网页 …… 4
2.3 Web浏览器与浏览Web …… 4
2.4 使用浏览器 …… 5
2.5 Web网站及其服务的实例 …… 5
2.6 Internet搜索 …… 6
2.7 小结 …… 6

第一部分 介绍网络互连

第3章 电话的普及 …… 8
3.1 引言 …… 8
3.2 通信服务 …… 8
3.3 拆销通信服务 …… 8
3.4 受限访问 …… 8
3.5 高昂的费用 …… 9
3.6 艰难的运营 …… 9
3.7 访问无处不在 …… 9
3.8 电话系统与Internet的关系 …… 10
第4章 过去的模拟技术时代 …… 11
4.1 引言 …… 11
4.2 声音，图画与模拟记录 …… 11
4.3 模拟电子设备 …… 12
4.4 许多电子设备都是模拟设备 …… 12
4.5 早期的模拟通信 …… 12
4.6 模拟简单但不精确 …… 12
4.7 在线路上传送模拟信号 …… 12
4.8 数字音乐 …… 13
4.9 数字革命 …… 13
4.10 计算机是数字式的 …… 14
4.11 数字录音 …… 14
4.12 使用数字显现模拟信号 …… 14
4.13 为什么选择数字音乐 …… 15
4.14 小结 …… 15
第5章 数字网络的过去与将来 …… 16
5.1 引言 …… 16
5.2 早先的数字时代 …… 16
5.3 电报是数字的 …… 16
5.4 莫尔斯码 …… 16
5.5 莫尔斯码中的字母与数字 …… 17
5.6 用户不会接触莫尔斯码 …… 17
5.7 事实上的即时通信 …… 18
5.8 速度是相对的 …… 18
5.9 数字电话 …… 18
5.10 与Internet的关系 …… 18
5.11 Internet中数据的二进制编码 …… 18
5.12 为何仅使用两个字符 …… 19
5.13 小结 …… 19
第6章 基本通信 …… 20
6.1 引言 …… 20
6.2 利用电话进行通信 …… 20
6.3 线路上的信号 …… 20
6.4 信息编码技术 …… 20
6.5 调制解调器及其双向通信 …… 21
6.6 数字信息的字符编码 …… 21
6.7 错误检测 …… 22
6.8 小结 …… 23
第7章 局域网时代的到来 …… 24
7.1 引言 …… 24

7.2 动机 …… 24
7.3 可交换介质 …… 24
7.4 计算机由集成电路组成 …… 24
7.5 将电路板插入计算机 …… 25
7.6 连接两台计算机 …… 25
7.7 局域网技术 …… 26
7.8 将计算机接入局域网 …… 27
7.9 局域网技术的重要性 …… 27
7.10 局域网与 Internet 的关系 …… 28

第二部分 Internet 简史

第 8 章 早期的 Internet …… 30
8.1 许多独立的网络 …… 30
8.2 数量激增的局域网 …… 30
8.3 有关局域网的几点事实 …… 30
8.4 局域网基本兼容的 …… 31
8.5 广域网技术 …… 31
8.6 少量广域网与大量局域网 …… 32
8.7 广域网与局域网互不兼容 …… 32
8.8 单一网络的愿望 …… 32
8.9 美国国防部拥有多个网络 …… 32
8.10 连接分散的计算机 …… 32
8.11 Internet 的出现 …… 33
8.12 ARPANET 主干网 …… 33
8.13 Internet 软件 …… 34
8.14 TCP/IP …… 34
8.15 开放式系统的影响 …… 34
8.16 开放式系统的必要性 …… 35
8.17 TCP/IP 联机文档 …… 35
8.18 军事部门采用 TCP/IP …… 36
8.19 小结 …… 36
8.20 个人见解 …… 36
第 9 章 难以置信的 20 年发展期 …… 37
9.1 引言 …… 37
9.2 传播 Internet 软件 …… 37
9.3 同时期计算机科学的支持 …… 37
9.4 Internet 适合 UNIX …… 38
9.5 美国军方的认可 …… 38
9.6 Internet 的规模在一年内扩大了一倍 …… 38
9.7 计算机科学部门 …… 39
9.8 研究生自愿付出他们的时间 …… 39
9.9 IAB 的演变 …… 39
9.10 IETF …… 40
9.11 Internet 规模一年内再次翻番 …… 40
9.12 Internet 促进科学发展 …… 40
9.13 NSF 的领导角色 …… 41
9.14 目标：涵盖所有科学和工程研究人员 …… 41
9.15 NSF 的做法 …… 41
9.16 NSFNET 主干网 …… 42
9.17 ANS 主干网 …… 42
9.18 Internet 呈指数级增长 …… 43
9.19 商业评估 …… 43
9.20 Internet 增长的尽头 …… 44
第 10 章 全球化的 Internet …… 45
10.1 引言 …… 45
10.2 早期的 ARPA 网络 …… 45
10.3 计算机间传送的电子邮件 …… 45
10.4 BITNET 与 FIDONET …… 45
10.5 欧洲的网络 …… 46
10.6 EBONE：欧洲的 Internet …… 47
10.7 主干网和 Internet 的层次结构 …… 47
10.8 各大洲的 Internet …… 48
10.9 1998 年后的 Internet 世界 …… 48
10.10 个人见解 …… 49
第 11 章 全球的信息基础设施 …… 50
11.1 引言 …… 50
11.2 现存的基础设施 …… 50
11.3 通信基础设施 …… 50
11.4 Internet 基础设施 …… 51
11.5 Internet 提供多样化的信息服务 …… 51
11.6 TCP/IP 提供通信机制 …… 51
11.7 个人见解 …… 52

第三部分 Internet 工作原理

第 12 章 包交换技术 …… 54
12.1 引言 …… 54
12.2 共享节省开支 …… 54
12.3 共享导致延迟 …… 54
12.4 共享线路 …… 54
12.5 可选的通道 …… 55
12.6 轮流共享 …… 55
12.7 包交换技术可以避免延迟 …… 55
12.8 必须标记每一个数据包 …… 56
12.9 计算机地址 …… 56
12.10 数据包的大小不都是一样的 …… 56
12.11 数据包传输似乎是即时的 …… 56

12.12 共享是自动的 …… 56
12.13 网络硬件处理共享 …… 57
12.14 许多设备都能使用包交换技术 …… 57
12.15 与 Internet 的关系 …… 57
12.16 小结 …… 57
第 13 章 Internet：网中网 …… 58
13.1 引言 …… 58
13.2 网络技术互不兼容 …… 58
13.3 解决不兼容问题 …… 58
13.4 两个基本概念 …… 58
13.5 利用计算机互联网络 …… 60
13.6 互联的计算机传递数据包 …… 60
13.7 连接网络的计算机称为路由器 …… 60
13.8 路由器是构建 Internet 大厦的基石 …… 60
13.9 Internet 包含多种网络 …… 61
13.10 路由器可以互联广域网和局域网 …… 61
13.11 Internet 的层次结构 …… 62
13.12 数据包传送的过程 …… 62
13.13 网络互联具有革命性 …… 63
13.14 小结 …… 63
第 14 章 ISP：宽带与无线接入技术 …… 64
14.1 引言 …… 64
14.2 ISP 及其资费 …… 64
14.3 用户连接形成"最后一英里问题" …… 64
14.4 租用专线费用昂贵 …… 65
14.5 拨号访问线路便宜但缓慢 …… 65
14.6 宽带提供高速连接 …… 65
14.7 重要概念：持续连接 …… 66
14.8 即时访问改变 Internet 使用方式 …… 66
14.9 现代技术提供便宜的专线接入服务 …… 67
14.10 无线访问的可用性 …… 68
14.11 Wi-Fi 与 3G 无线技术 …… 68
14.12 个人见解 …… 69
第 15 章 IP：建立虚拟网络的软件 …… 70
15.1 引言 …… 70
15.2 协议：通信协定 …… 70
15.3 基本的功能协议：网际协议 …… 70
15.4 每台机器上的 IP 协议软件 …… 70
15.5 Internet 数据包称为数据报 …… 70
15.6 巨型网络的设想 …… 71
15.7 Internet 的内部结构 …… 71
15.8 数据报装在数据包中传输 …… 72
15.9 每一台计算机分配一个唯一的地址 …… 72
15.10 Internet 地址 …… 73
15.11 奇怪的 IP 地址语法 …… 73
15.12 IP 地址不是任意的 …… 73
15.13 示例网络上的传输过程 …… 73
15.14 你的网络连接有多快 …… 74
15.15 小结 …… 74
第 16 章 TCP：使通信可靠的软件 …… 75
16.1 引言 …… 75
16.2 包交换系统会发生过载 …… 75
16.3 TCP 协议帮助 IP 协议确保数据传输 …… 76
16.4 TCP 协议提供计算机程序间的连接 …… 76
16.5 恢复丢失的数据报 …… 77
16.6 TCP 协议的自动重传 …… 77
16.7 TCP 协议和 IP 协议协同工作 …… 77
16.8 小结 …… 78
第 17 章 客户机-服务器-分布式计算 …… 79
17.1 引言 …… 79
17.2 大型计算机利用网络进行输入输出 …… 79
17.3 小型计算机利用网络进行交互 …… 79
17.4 Internet 上的分布式计算 …… 80
17.5 通过一个范例解释所有分布式计算 …… 80
17.6 通信程序分为客户机和服务器 …… 81
17.7 服务器必须一直运行 …… 81
17.8 小结 …… 81
第 18 章 计算机的命名 …… 82
18.1 引言 …… 82
18.2 比起数字来人们更喜欢用名字 …… 82
18.3 为计算机命名既困难又有趣 …… 82
18.4 计算机的名字必须唯一 …… 83
18.5 计算机名字的后缀 …… 83
18.6 名字由多部分组成 …… 84
18.7 美国之外的计算机域名 …… 84
18.8 将计算机的名字转换成对应的 IP 地址 …… 85
18.9 域名系统的工作方式类似于电话系统的目录帮助服务 …… 85
18.10 计算机名字的查询是自动的 …… 85

18.11 IP 地址与域名并不相关 …… 86
18.12 小结 …… 87
第 19 章 NAT：共享 Internet 连接 …… 88
19.1 引言 …… 88
19.2 大容量与多台计算机 …… 88
19.3 共享一个 IP 地址是可能的 …… 88
19.4 实现共享连接的设备称为 NAT 盒 …… 88
19.5 NAT 盒的行为与小型 ISP 类似 …… 89
19.6 NAT 改变每个数据报中的地址 …… 89
19.7 计算机软件可以实现 NAT 功能 …… 90
19.8 NAT 可以用于无线网络 …… 90
19.9 小结 …… 90
第 20 章 为何 Internet 运行良好 …… 91
20.1 引言 …… 91
20.2 Internet 运行良好 …… 91
20.3 IP 协议提供灵活性 …… 91
20.4 TCP 协议提供可靠性 …… 92
20.5 TCP/IP 协议软件的开发注重高效性 …… 92
20.6 TCP/IP 协议的研究着重于实际结果 …… 92
20.7 成功的法则 …… 93
20.8 小结 …… 93

第四部分 可以使用的 Internet 服务

第 21 章 电子邮件 …… 96
21.1 引言 …… 96
21.2 功能描述 …… 96
21.3 全世界的福音 …… 96
21.4 电子邮件用户都拥有一个邮箱 …… 96
21.5 发送电子邮件消息 …… 97
21.6 电子邮件到达时的通知 …… 97
21.7 读取电子邮件消息 …… 97
21.8 使用浏览器发送和接收电子邮件 …… 97
21.9 电子邮件类似于办公室间的备忘录 …… 97
21.10 电子邮件软件填充邮件头信息 …… 98
21.11 电子邮件服务如何工作 …… 98
21.12 在个人计算机上使用电子邮件 …… 99
21.13 邮箱地址的格式 …… 99
21.14 编写使电子邮件服务变得友好 …… 100
21.15 别名支持任意缩写 …… 100
21.16 在一个计算机系统内所有用户共享的别名 …… 100
21.17 向多个收件人发送邮件 …… 101
21.18 邮件列表：指定多个收件人的别名 …… 101
21.19 公共邮件列表和邮件检查程序 …… 101
21.20 与非 Internet 网站交换电子邮件 …… 102
21.21 利用电子邮件访问其他服务 …… 102
21.22 速度、可靠性和期望 …… 103
21.23 电子邮件服务的影响和意义 …… 103
21.24 加入邮件列表的约定 …… 104
第 22 章 电子公告板服务（新闻组） …… 105
22.1 引言 …… 105
22.2 功能描述 …… 105
22.3 具有多样话题的电子公告板 …… 105
22.4 网络新闻 …… 106
22.5 公告板的种类 …… 106
22.6 获取网络新闻及阅读文章的软件 …… 107
22.7 用户眼中的网络新闻 …… 107
22.8 查阅新文章 …… 107
22.9 阅读网络新闻 …… 108
22.10 提交文章 …… 108
22.11 受限新闻组 …… 108
22.12 网络新闻的规模 …… 108
22.13 新闻组和邮件列表的影响及意义 …… 109
22.14 参与讨论的建议和约定 …… 109
22.15 小结 …… 110
第 23 章 浏览万维网 …… 111
23.1 引言 …… 111
23.2 功能描述 …… 111
23.3 浏览与信息检索 …… 111
23.4 早期的浏览服务采用菜单驱动方式 …… 112
23.5 菜单项可以指向另一台计算机 …… 112
23.6 浏览器如何工作 …… 112
23.7 点击式界面的示例 …… 113
23.8 菜单项与文本相结合 …… 113
23.9 集成式链接的重要性 …… 114
23.10 在文本中嵌入的链接称为超文本 …… 115
23.11 多媒体 …… 116
23.12 文本中可以嵌入视频和音频引用 …… 116
23.13 万维网 …… 117

23.14 用于访问 Web 的浏览器软件 …… 117
23.15 超媒体显示的示例 …… 118
23.16 浏览器的控制 …… 118
23.17 外部引用 …… 119
23.18 记录信息位置 …… 119
23.19 书签/收藏夹 …… 120
23.20 万维网如何工作 …… 120
23.21 URL 告诉浏览器所要联系的计算机 …… 120
23.22 URL 告诉浏览器所要联系的服务器 …… 121
23.23 URL 中"www"的使用 …… 121
23.24 浏览器提供对多种服务的访问 …… 121
23.25 浏览器程序的内部 …… 122
23.26 小结 …… 122
23.27 对超媒体浏览的评论 …… 123
第 24 章 万维网文档（HTML） …… 124
24.1 引言 …… 124
24.2 显示硬件的改变 …… 124
24.3 浏览器解释并显示 Web 文档 …… 124
24.4 Web 所用数据的结果 …… 125
24.5 HTML：用于 Web 文档的语言 …… 125
24.6 Web 页面中的指令能够控制输出 …… 125
24.7 Web 页面分为两个主要部分 …… 126
24.8 缩排使 HTML 易读 …… 126
24.9 在 Web 页面的正文中加入文本 …… 127
24.10 缩排使段落更易被发现 …… 127
24.11 一个 Web 页面能够链接到另一个 Web 页面 …… 128
24.12 HTML 支持已编号的和未编号的列表 …… 128
24.13 Web 页面中的图像是数字的 …… 129
24.14 HTML 支持在 Web 页面中加入图像 …… 130
24.15 文本可以出现在图像周围 …… 130
24.16 图像能够链接到另一个 Web 页面 …… 131
24.17 有些浏览器能放大或缩小图像 …… 131
24.18 背景控制 …… 132
24.19 HTML 的其他特色 …… 132
24.20 HTML 的重要性 …… 132
24.21 使用 GUI 工具创建 Web 页面 …… 132
24.22 小结 …… 133
第 25 章 高级网页技术（表单、框架、插件、Java、JavaScript 脚本和 Flash 动画） …… 134
25.1 引言 …… 134
25.2 传统 Web 采用静态页面 …… 134
25.3 服务器如何存储静态页面 …… 134
25.4 一次提取一项条目 …… 135
25.5 传统的 Web 页面使用整个屏幕 …… 136
25.6 Web 页面能改变屏幕的局部显示 …… 136
25.7 Web、广告和框架 …… 136
25.8 弹出窗口与弹出窗口阻止程序 …… 137
25.9 静态页面的缺点 …… 137
25.10 控制浏览处理数据的过程 …… 139
25.11 使用插件实现更多功能 …… 139
25.12 服务器可以按需生成页面 …… 139
25.13 服务器端脚本如何工作 …… 139
25.14 程序员创建服务器端脚本 …… 140
25.15 个性化 Web 页面 …… 140
25.16 个性化广告 …… 141
25.17 Web 页面也可以交互 …… 141
25.18 脚技术 …… 141
25.19 cookie …… 141
25.20 应该禁用 cookie 吗 …… 142
25.21 Web 页面中也可以播放动画 …… 142
25.22 活动页面功能更强大 …… 143
25.23 Java 是一种活动页面技术 …… 143
25.24 JavaScript 脚本也是活动页面技术 …… 144
25.25 Flash 动画和 Real 流媒体技术 …… 144
25.26 高级 Web 技术的重要作用 …… 144
第 26 章 群组和个人网页（维基、博客） …… 145
26.1 引言 …… 145
26.2 论坛的缺点 …… 145
26.3 共享页面 …… 145
26.4 共享页面就是维基 …… 146
26.5 共享与共识 …… 146
26.6 维基的缺点 …… 146
26.7 维基百科是一个试验 …… 146
26.8 维基百科是否可信 …… 147
26.9 个人日记的公开发布 …… 147
26.10 个人见解 …… 147

第 27 章 网页搜索（搜索引擎） …… 146
27.1 引言 …… 146
27.2 搜索引擎的功能 …… 148
27.3 浏览还是搜索 …… 149
27.4 搜索引擎可以更好地帮助用户入门 …… 149
27.5 搜索工具可以使用户避免信息丢失 …… 149
27.6 自动搜索服务是怎样运行的 …… 149
27.7 抓先收集信息 …… 150
27.8 现代系统可以搜索页面内容 …… 150
27.9 网页搜索服务的用户界面 …… 150
27.10 搜索引擎怎样返回结果 …… 151
27.11 自动搜索使用字符串匹配方法 …… 151
27.12 字符串匹配方法的优劣 …… 151
27.13 支持多关键字的自动搜索 …… 152
27.14 高级服务可以完成更复杂的搜索 …… 152
27.15 个性化的搜索结果 …… 152
27.16 内容搜索的更多细节 …… 153
27.17 搜索以及有关问题的 …… 154
27.18 广告为搜索服务付账 …… 154
27.19 自动搜索服务的例子 …… 154
27.20 自动网页搜索的重要性 …… 155
第 28 章 文字、音频和视频通信（IM 和 VoIP） …… 156
28.1 引言 …… 156
28.2 即时消息提供文字交流机会 …… 156
28.3 音频和视频功能 …… 156
28.4 音频和视频需要特殊的设备 …… 157
28.5 音频剪辑如同一张 CD …… 157
28.6 实时意味着没有延迟 …… 157
28.7 Internet 上的实时音频 …… 157
28.8 Internet 的广播电台 …… 158
28.9 实时音频传输也称为网上广播 …… 158
28.10 Internet 电话服务会成为现实 …… 159
28.11 Internet 电话就是所谓的 VoIP …… 159
28.12 音频电话会议 …… 159
28.13 协同文档标记服务 …… 160
28.14 对文档进行标记 …… 161
28.15 参与者共同讨论并标记文档 …… 161
28.16 视频电话会议 …… 162
28.17 小组视频电话会议 …… 162
28.18 同时提供音频、视频、白板和即时消息服务 …… 163
28.19 小结 …… 163
28.20 个人见解 …… 163
第 29 章 传真、文件传输和文件共享（FTP） …… 164
29.1 引言 …… 164
29.2 发送传真 …… 164
29.3 可以通过 Internet 拷贝文件 …… 164
29.4 文件中的数据 …… 164
29.5 拷贝文件 …… 164
29.6 在浏览器中使用 FTP …… 165
29.7 通过 FTP 浏览目录内容 …… 165
29.8 通过 FTP 上传文件 …… 165
29.9 通过 FTP 传输文件必须经过授权 …… 166
29.10 FTP 工作原理 …… 166
29.11 FTP 的影响和重要性 …… 166
29.12 对等方式的文件共享 …… 167
29.13 小结 …… 167
第 30 章 远程登录和远程桌面（TELNET） …… 168
30.1 引言 …… 168
30.2 早期计算机使用字符界面 …… 168
30.3 分时系统需要为每个用户分配标识 …… 168
30.4 远程登录和本地登录几乎没有差别 …… 168
30.5 远程登录工作原理 …… 169
30.6 远程登录过程中的转义问题 …… 169
30.7 Internet 远程登录的标准是 TELNET …… 169
30.8 远程访问也可以提供桌面 …… 170
30.9 远程桌面的运行机制 …… 170
30.10 远程登录和远程桌面的重要性 …… 170
30.11 通用性让远程登录和远程桌面更强大 …… 171
30.12 远程访问可用于多种类型计算机 …… 171
30.13 远程访问的一些异常结果 …… 171
30.14 小结 …… 172
第 31 章 安全通信技术 …… 173
31.1 引言 …… 173

31.2 Internet 并不安全 …… 173
31.3 缺乏安全性是致命的缺陷 …… 173
31.4 认证和保密是基本问题 …… 173
31.5 数据可以被篡改 …… 174
31.6 编码可以保证消息的私密性 …… 174
31.7 计算机的密码使用了数学 …… 174
31.8 没有绝对安全的网络 …… 174
31.9 加密保障了电子邮件的私密性 …… 175
31.10 加密软件需要密钥 …… 175
31.11 两把密钥意味着不需要预先信任关系 …… 175
31.12 安全电子邮件的实际应用 …… 176
31.13 安全的无线网络 …… 176
31.14 防火墙可以阻止不想要的数据包 …… 177
31.15 防火墙可以对数据包进行过滤 …… 177
31.16 防火墙可以防止特洛伊木马 …… 177
31.17 住宅和个人防火墙 …… 178
31.18 有一类程序可以用来检测黑客入侵 …… 178
31.19 服务可能被拒绝 …… 178
31.20 小结 …… 179
第 32 章 远程安全接入（VPN） …… 180
32.1 引言 …… 180
32.2 组织机构中的员工享有特殊的权限 …… 180
32.3 外出的员工不再有特殊的权限 …… 180
32.4 远程用户没有特殊的权限 …… 180
32.5 租用专线可以保证通信安全 …… 181
32.6 标准的 Internet 连接更便宜 …… 181
32.7 能否兼顾二者之长 …… 181
32.8 虚拟专用网是一个解决方案 …… 181
32.9 VPN 怎样工作 …… 182
32.10 所谓的"直接连接" …… 182
32.11 VPN 的重要性 …… 183
第 33 章 Internet 经济和电子商务 …… 184
33.1 引言 …… 184
33.2 谁为 Internet 买单 …… 184
33.3 电子商务是一块大蛋糕 …… 184
33.4 安全技术使得电子商务成为可能 …… 185
33.5 安全套接字 …… 185
33.6 公钥加密技术提供了认证手段 …… 186
33.7 数字签名 …… 186
33.8 证书中的公钥 …… 186
33.9 什么是电子货币 …… 186
33.10 电子现金还未广泛应用 …… 187
33.11 企业和电子商务 …… 187
33.12 关于税收和网络内容中立性 …… 187
第 34 章 全球数字化图书馆 …… 189
34.1 引言 …… 189
34.2 各种各样的服务 …… 189
34.3 新服务总是有规律地出现 …… 189
34.4 灵活性可以适应变化 …… 189
34.5 数字化图书馆 …… 190
34.6 卡片目录和搜索工具 …… 190
34.7 Internet 上的服务可以融合 …… 190
34.8 Dewey 先生，你在哪 …… 191
34.9 数字化图书馆中的信息 …… 191
34.10 Internet 是什么 …… 192
34.11 个人见解 …… 192
Internet 术语表 …… 193

第 1 章　Internet 时代已经到来

1.1　世界已经改变

一场革命发生了。它静悄悄地开始，现在已经席卷了大部分世界。在普通的一天里会发生下面这些事情：

- 一对年轻夫妇用计算机设计从俄亥俄州自己家中到加州旧金山的朋友家里的一次道路旅行的路线。在数秒之内，他们就在地图上得到了最短路径的描述以及各段距离和每个岔口的详细资料。
- 在德州休斯敦，一名投资经纪人坐在一台个人计算机前操作若干计算机程序，该程序可以获取纽约证券交易所的当前的股票价格信息。当看见这张价格列表后，他买进了两只股票，同时又抛售了另一只股票。
- 在伊利诺斯州的芝加哥市，小学生们通过计算机网络阅读《纽约时报》。
- 一位居住在马里兰的母亲在计算机上查看女儿所居住的北卡罗来纳的天气情况。当得知那里下过一场暴雨后，她给女儿发了一封慰问信。
- 在华盛顿的西雅图，一位十多岁的青少年通过计算机试听音乐。过了一会儿，他用计算机购买并下载了这首歌。
- 在马萨诸塞州的波士顿市，一位祖父准备去看望自己的孙子。他在计算机上查询航班信息、预订旅馆房间，并购买机票。
- 一家公司的主管正在开会，其中一名主管身在纽约，另一名在佛罗里达，还有一位在科罗拉多州。他们三个人都坐在带摄像头和麦克风的计算机前，他们听见对方的声音并且在显示屏上看到对方的头像。
- 下午 6 点，位于位治亚州亚特兰大市的一个计算机程序把公司每日销售额发送给在英国伦敦的分支机构。
- 一名台湾的中学生在计算机上观看夏威夷一所大学的校园游——包括同学们步行穿过布满在微风中摇曳的棕榈树的校园的现场视频录像。不久之后，这名学生用计算机给一个正在澳大利亚旅游的亲戚发送了一个传真。

上述这些事件都有什么共同点呢？我们容易看出，在每件事情中，人们都用到了 Internet——这个交流系统正在让我们的工作和娱乐方式发生革命性的改变。

如果你还没有接触过 Internet，那不久你就会感受到它的。先让我们看看一些统计数据：

- Internet 至今已经拥有遍及超过 205 个国家（地球上所有有人口居住的地区）的成百上千万的用户。
- 美国两年制和四年制的学院，综合性大学都接入了 Internet，公立学校也不例外。
- 美军使用 Internet 技术超过 20 年。早在 20 世纪 80 年代初的“沙漠风暴”行动中 Internet 就在军事行动中发挥了重要作用。
- 科学家自 1980 年就开始使用 Internet。
- 通过 Internet 可以访问美国总统和白宫，以及设在许多国家的其他政府机构。

1.2　数字不能说明问题

评估Internet重要性的最常用方式是统计接入Internet的计算机总数，但是传统的计算机连接只能说明问题的一部分。Internet已经触及到航海、航空以及地面运输领域。一些私人公司通过电话系统和有线电视系统提供访问Internet的服务，从而使得任何装有电话和有线电视线缆的家庭或办公室都有可能访问Internet。

为了评估Internet的影响力，有人可能会问："它到底影响了什么？"答案是："几乎无所不包。"所以，这个问题就变成了下面这个样子：

Internet时代已经到来，它又是如何影响你的呢？

1.3　学习Internet

本书从广义上回答了"什么是Internet"的问题，主要介绍了计算机网络的起源以及如何用它解决日常问题。本书重点介绍了Internet提供的各种服务，帮助读者理解这些服务的重要性。更加重要的是，本书解决了一些疑难问题，帮助读者明白Internet是如何运作的。

学习Internet不是一个下午就能完成的事——对它的学习永远都不能停下来。因为Internet时时刻刻都在发生着变化。Internet就像一个报摊，每当新信息出现时，旧的信息就会被取代。你每次到报摊询问或到Internet浏览时，都会找到一些新的东西。

当然，Internet中的信息更新速度远远大于一个传统的报摊。事实上，由于Internet上的信息来自于计算机和自动化系统，所以信息能够立即发生变化。例如，有人在一分钟内两次查询天气状况，那么这两次查询反馈的结果有可能不同，这是因为计算机在不停地测量天气情况并能即时更改气象报告。

除了类似于报摊，Internet还与图书馆相似。因为它也有帮助检索信息的各种工具。从历史上延续下来，图书馆一般都会给读者提供卡片式目录和检索台，而Internet同样能够提供类似的服务，以电子方式帮助读者找到所需信息。

1.4　了解Internet的概貌

由于Internet一直处于快速发展中，所以很难掌握它的所有细节。例如，没人能知道自己感兴趣的数据的具体位置，也没人能知道购买最低价商品的有效途径。更为重要的是，由于新的应用在不断出现，没人能对所有服务作出完整的描述。由于各种类型的个人计算机及软件程序间差异，同一细节不可能适用于所有计算机。

为了避免读者被过多的细节问题所困扰，我们将侧重讲解Internet技术的基础知识。我们并不重点介绍如何使用一台特定的计算机、一款特定的软件或者一项特定的Internet服务，而是要探讨Internet基本的工作机制以及各种服务如何使用这一基本机制。从本质上讲，我们将介绍Internet的各项功能。

了解Internet的功能使阅读计算机手册和使用Internet更为容易。尤其是，大部分计算机手册仅仅是具体说明了如何完成一项任务的细节，而并不告诉人们为什么需要这样做，因此，初学者常常发现自己很难按照这些细节来做。了解Internet的工作原理和每项服务的目的可以帮助人们真正明白这些细节的作用。

1.5　术语和技术

作为一项复杂的技术，Internet在其领域已经产生了一套令人望而却步的术语。本书使用类

比和举例的方法清楚地解释了Internet技术。本书将阐明各种技术如何结合在一起，在避开细节的同时奠定牢固基础。本书还将讨论Internet提供的各项服务，并解释这些服务的灵活性和用法。

更为重要的是，本书介绍了计算机网络和Internet领域的相关术语。我们不是简单地提供一张术语表，而是从历史的角度，即从介绍通信系统的演化过程入手，给出每个术语的定义。例如，本书前几章解释了数字信息与模拟信息的区别，然而这几章并没有以计算机网络为例，而是将这些术语与日常生活经历联系起来进行说明。

1.6 Internet的发展及其适应性

人们对Internet的神秘感部分源于它的巨大成功。随着Internet的不断发展，其他许多提供相同服务的尝试都以失败告终。同时，因为Internet能够适应技术上和策略上的变化，所以它得以不断壮大。我们将研究Internet技术运行如此之好的原因以及它又是如何适应各种各样的变化的。

Internet的另一个令人吃惊的地方是它的不可思议的成长。我们将目睹Internet如何持续发展以及这样发展的结果。

1.7 Internet的影响

Internet最重要的方面也许就是它对社会的影响。Internet曾经被局限于少数科学家使用，但是很快便普及开来。它涵盖的范围涉及政府机构、商业、学校以及全世界的家庭。我们将说明Internet是如何影响人们的生活的，以及我们期待将来会怎样。总之，本书的其余部分将展示Internet是什么以及它能对您起什么作用。

1.8 本书的组织结构

本书由四部分组成。我们将在第2章对各类网站进行一次简单的了解，之后，第一部分（第3章～第7章）将介绍通信系统的概念和相关术语。如果你已经了解了数字通信技术和模拟通信技术，普通服务以及二进制编码方式，那么你就可以略过这部分章节。第二部分（第8章～第12章）回顾了Internet的发展史及其令人难以置信的成长历程。本部分展示了数字革命进展的速度，给出了帮助你评价底层架构设计重要性的背景。第三部分（第13章～第20章）描述了基本的Internet技术及功能。本部分考察了Internet硬件设施的组织方式，以及软件如何提供通信服务。读者一定要理解这部分的内容，因为它是后续章节的基础。最后一部分描述了当前可用的各种Internet服务。对于每种服务，本部分都阐释了它的运作方式和用法。

1.9 个人见解

我至今仍然记得，在几年前，有位同事直截了当地问我："Internet是什么？"当时，我从事Internet研究已经许多年了，也写过一本比较受欢迎的介绍Internet及其底层设计原理的大学教材，我了解构成Internet的软硬件系统的许多细节，懂得计算机之间如何相连，也知道通信过程的各种细节。我还认识大多数正在从事Internet技术改进的研究人员。最让我困惑的是：问这个问题的人已经知道基本的技术细节，而且也看过我写的教材，那么我该怎么回答呢？

当我正在思考如何回答这个问题的时候，我的同事以为我误解了他的意思，于是对我说："我不想了解计算机和各种线路，我的意思是，从广义上讲，Internet到底是什么？它将来会变成什么样子？你注意到Internet发生的变化了吗？十年后谁会用到它？人们又会个它做什么呢？"

这些问题十分重要，因为它们指出了一个重大的转变。在Internet发展的初期，大部分用户都是那些建造它的专家，然而现在，Internet早已超出了仅作为研究的原态，成为一种强大的网络工具，几乎每个人都在使用它。Internet现在的用途是当时所有专家都始料不及的。

第 2 章　开始吧：亲身体验

2.1　引言

本章中的资料从宽泛和概念的角度介绍了 Internet 的各种功能。这些资料解释了 Internet 的用途以及其底层技术的运作方式，但并没有注重细节的诠释。例如，本章既不会列出 Internet 中藏有哪些东西的目录，也不会讨论如何使用一台特定的计算机或者一套特定的软件。因此，在没有直接接触计算机的情况下，我们也能够理解本章的内容。

尽管本章主要强调的是概念，但是我们仍然鼓励具有 Internet 访问条件的读者在阅读的同时使用 Internet。亲身体验既能使用户直观地了解 Internet，也能使用户熟悉 Internet 环境。这些都能增强 Internet 相关知识的学习效果，这与亲自驾车能帮助更好的了解公路状况及这些公路能够到达那些地方有异曲同工之妙。为了帮助读者尽快入门，本章提出了一些可行性建议。本章描述了 Web 浏览器的概念——一个计算机程序，给出了一些可以访问的网站的示例。第 23 章将次讨论了浏览器的话题，并且从更加详细的角度解释了它是如何运作的。已经熟悉 Internet 的读者不需要花费时间阅读本章，可以直接跳到下一章。

2.2　Web：网站与网页

虽然为 Internet 设计了许多应用，但是有一项已经发展成熟的应用能通过一个接口整合其他应用的最好特性。这项应用就是大家熟知的万维网（World Wide Web，简称 Web）。

大部分的大机构或团体都有自己的 Web 网站，一些小机构或团体亦如此。我们把一个机构的 Web 网站看作联系它的平台，该网站包含了有关该机构的所有信息。这些信息对于 Internet 用户来说都是可用的。例如，一个公司的 Web 网站可能包含该公司产品及其价格、订购说明的目录列表，公司雇员的电话号码、职责的列表，或者公司招聘的相关信息。除了公司之外，像学校这样的研究机构以及政府部门也都有自己的 Web 网站。最后要说明点，有些人开发自己的 Web 网站，这些网站主要包含的是一些个人信息。

为了避免用户被大量的信息淹没，Web 网站的所有内容被分成不同的"页面"（page），我们称之为"网页"。一般而言，Internet 用户一次浏览一个网页的内容。因此，网页通常都被设计成刚好适合用户的显示屏大小。当用户打开新的网页时，前一网页的内容不再显示出来。要再次浏览前一网页，用户必须明确返回到前一页[1]。

2.3　Web 浏览器与浏览 Web

访问万维网需要一台接入 Internet 的计算机和运行在该计算机上的特定软件。在后续章节中，我们将知道这里需要的是两类软件：Internet 通信软件和应用软件。通信软件讲的是"Internet 语言"，这使计算机之间的通信成为可能；应用软件处理用户之间的各种交互。当用户发出一个请求后，应用软件通过 Internet 访问用户需要的信息，然后在屏幕上显示结果以供用户浏览。

---

[1] 用于显示信息的应用程序可以在用户的屏幕上创建多个窗口以便于用户同时浏览几个网页。

一个用于访问万维网的应用程序就是Web浏览器（Web Browser），人们把使用这样一个程序称作进浏览（browsing）Web。尽管已经有好几种浏览器问世，但是只有其中两种支配着浏览器市场：Mozilla的火狐（Firefox）浏览器和Microsoft的Internet Explorer，它们仅仅在次要的细节上有些不同。值得注意的是，这两种浏览器都是可以免费使用的。

2.4 使用浏览器

浏览Web的方式十分简单。首先，用户需要把自己的计算机接入Internet并打开一个浏览器。浏览器开始运行后，它会在屏幕上创建一个窗口，该窗口主要包括一个用于显示Web网页的区域（除非用户进行过一些特定的设置，否则浏览器也很可能显示伴随网页弹出的广告栏）。在浏览器窗口顶部附近有一个小区域，它包含一些控制模块，这些模块供用户调用以便指定一个网页。浏览器在控制区有一个较小的被标记为"*location*"的文本框，此文本框用于指定一个特定的Web网站地址。每个Web网站都分配有一个标识符串，称作URL（后面的章节将解释这个术语，而现在没有必要解释它）。要访问一个Web网站，用户只要把鼠标移动到地址栏，输入该网站的URL，然后按回车键。之后，浏览器就会连接这个网站并显示它的主网页。该网页可以包含文本、图片和音频信息。

一旦一个网页显示在屏幕上，转到另一个网页就比较容易。网页上的一些条目是指向其他网页的链接。作为链接的文本往往以不同的颜色并加下划线显示。虽然网页上的某些图片也被作为链接，但是它们可能并没有明显的表示出来。要打开一个链接，用户使用鼠标把指针移动到这个条目的上方，然后单击它。之后浏览器就自动通过Internet获取该链接对应的网页，最后代替现有网页显示新网页。因此，浏览Web主要由两步组成：首先，输入一个指定网站的URL；然后，跟随链接从一个网页转换到另一个。

2.5 Web网站及其服务的实例

一些例子有助于展示Web中不同类型的可用信息。文中所选的实例并不意味着它们是必选的网站，这仅仅是给出了一个范围，列出了多种可用信息。如果要访问这些网站之一，输入其对应的URL即可。

1. 在线报纸

纽约时报（New York Times）是世界上最著名的报纸之一，它的官方网站包含了当天报纸的各类文章和股票市场的各种信息。该网站的URL是：

www.nytimes.com

2. 气象信息

气象频道提供在线的气象信息和天气预报服务。它的URL是：

www.weather.com

3. 驾驶导航

有几家Web网站提供详细的驾驶导航服务，包括从一个地点到另一个地点的行车路线、行车距离和交路口信息。例如，MapQuest服务可以在下面的网址中找到：

www.mapquest.com

4. 卫星地图

一个拥有大量地球表面的卫星和航空地图的Internet网站。你应该能够看到你附近地区的卫星地图。其URL是：

www.terraserver.com

5. 广播站

许多广播站通过Internet发送音频信息，在下面的链接指向的页面中可以找到根据音乐文件格式分类的音乐目录列表，其URL是：

www.xmradio.com

6. 在线时尚购物中心

每个大城市都有一个容纳大量著名品牌商店的时尚购物中心。一个在线版本可以在下面的网址找到：

www.fashionmall.com

7. 零售书店

亚马逊（Amazon）是最著名的在线零售商之一。起初，它仅仅是一个在线书店，现在已经提供了许多其他商品。其URL是：

www.amazon.com

8. 股票信息

纽约证券交易所网站除了提供受关注的股票索引信息和股价信息外，还提供整个投资社区的信息。它的URL地址如下：

www.nyse.com

9. 音乐剪辑

第28章解释了音频信息是如何在Internet上传播的，并对MP3作为一个流行的音乐编码格式进行讲解。MP3格式的音乐和别的音频信息能够在下面的网站上找到：

www.itunes.com

10. 电子邮资

购买电子邮资并用它投寄传统的信件是完全可能的。本质上，像邮资计算器一样，电子邮资操作方式代替了传统的邮票——当支付完邮资后，客户打印一张邮局认可的标签，以此代替邮票。该网站的URL地址是：

www.e-stamp.com

11. 讽刺

许多网站都有一些调侃讽刺的内容。一则讽刺微软的资料看起来像一个平面游戏，它可以在下面的网站找到：

www.ms-monopoly.com

2.6　Internet搜索

如本书第27章所述，Internet搜索是一项提供用户找到所需信息条目或主题所在网页的网络服务。如下网址可以找到这样的服务：

www.google.com

2.7　小结

本章简要地提及了最受欢迎的Internet服务——万维网。我们了解到一个Web网站的信息被分成若干网页，用户使用浏览器（一种计算机程序）一次只能浏览一个网页。每个网页都可以包含文本和图片，也可以连接到音频或视频。我们还了解到各种各样的信息都可以通过Web进行访问。第21章～第26章更详细地说明了Web技术，并解释了Internet是如何提供这些服务的。

第一部分　介绍网络互连

前 Internet 时代

通信系统概念
与术语的简单介绍

第3章　电话的普及

3.1　引言

本章将介绍普遍服务（universal service）的概念。文中用人们熟悉的例子说明普遍服务的设想如何影响人们对通信服务的认识，解释为什么Internet在普及之后能够成为人们日常生活不可或缺的一部分。

3.2　通信服务

Internet是一种通信技术，就像先前的电话一样，Internet使人们有可能用新的方式进行通信。然而，就像生活在一百年前的人们对电话通信的感觉一样，对生活在现在的人们来说，数字通信是一件新鲜的事物。因此，电话服务的发展史可以直接用于理解Internet，人们可以从中得到更多的启示。

3.3　推销通信服务

为了便于理解一项新技术是如何渗透到社会当中去的，先让我们的思路回溯到近一个世纪以前，把你自己想像成一名在美国一个普通城镇推销电话服务的销售员。

从各个方面考虑，你所面对的那个经济时代都充满了希望，激动与乐观充斥着整个工业界。毕竟，当时的社会正经历一场工业革命，你会发现无论在哪个地方机械作业都已经替代了手工劳动，蒸汽机成为取代水轮机和畜力的动力源；部分行业正开始使用以汽油驱动的发动机；工厂生产出的商品比以往任何时候都要多得多。

当然，一个正在崛起的电话销售员几乎或者根本没有使用电话的经验。实际上，在推销电话服务之前，他们可能只见过几次使用电话的演示。

可以想像，当你走进一家小公司，向老板介绍电话服务时，你会对他说些什么呢？也许，你会说公司需要电话，因为电话使顾客订货更容易；或者你可能会说，雇员们可以通过电话与供应商洽谈、订购原材料或跟查那些货物没有按期抵达。也许，你会问老板，是否他要出去和其他公司老板共进午餐并紧接着指出，如果使用电话的话，可以在几十秒内把午餐安排好。你可能还会说电话是很容易使用的，甚至，你可以讲到更严重的方面，如果公司失火，打电话可以立即通知消防队，电话可用于保护公司财产和员工的生命安全。

那么，对于你的电话推销策略，公司的老板们会如何反应呢？可能一些人比较感兴趣，但是大多数还是持怀疑态度的。一些人爱听，而其他人觉得反感。尽管有些人会认为这种想法不错，但是多数人还是只会付之一笑。一部分人想要亲自做计业务上的操作方法，但是大部分人更愿意保持现状。一些人想要装上电话，仅仅因为电话是一件新鲜事物，他们也想借此提高一下自己的地位。不管推销员说什么，大多数老板还是认为，即使没有电话，公司的业务也会照常开展。

3.4　受限访问

没有用过电话就去销售电话可能会很困难。为了使销售工作容易些，我们假设你在拥有电

话的社会中成长起来，在你的生活中也经常使用电话。再假设，你又回到一百年前，去尝试销售电话服务，你也许会认为，说服人们购买电话服务是一件比较容易的事情，因为你自己知道怎么使用电话。然而，实际情况并非如此，你将对你所遇到的事情感到惊讶。

当销售电话服务的时候，遇到的第一件让你感到意外的事就是一个世纪前的电话服务跟现代电话服务的运营方式不一样，那时的电话服务是本地（local）服务，各城镇或乡村都是独立决定何时布线、何时雇用接线员以及何时建立电话服务。更重要的是，各城镇都选择满足自身需要和预算的电话技术。结果，尽管存在许多电话系统，但是它们之间互不兼容，也就是说，在两个城镇之间布线并不能保证两个电话系统协调工作。从商业角度看，即便一家公司安装了电话，也不能用它从国内的其他地区订购货物。你很快就会发现：

> 各个城镇之间相互独立的本地电话服务限制了电话的用途。

3.5 高昂的费用

当你试着在近一个世纪之前推销电话服务的时候，第二件令你感到意外的事就是，即使有电话可用，它的费用也很昂贵，普通家庭没有能力支付家用电话的费用。除了购买电话本身之外，许多电话公司还向用户收取实实在在的安装费。某条街道上的第一个电话用户必须支付从电话公司到该街道之间的布线费用，而这之后的电话用户只需要替付接到他们家的线路费用。因此，要想在某个小区找寻第一个用户往往比找其他用户困难得多，更为重要的是，对大部分生活在乡村的人们来说，享受电话服务根本就是不可能的。

在多次尝试向个人用户推销电话服务失败后，你向你的老板汇报的结论是：

> 在电话安装的费用降到普通家庭所能承受的范围之前，电话服务是不可行的。

3.6 艰难的过渡

在没有电话的社会里，说服一家公司安装一部电话似乎是不可能的事情，如果公司不能打电话给远在其他地区的供应商，而且本地用户也无法方便地使用电话，那么安装电话就没有什么经济价值。事实上，仔细比较分析我们所享受的电话服务和大约一个世纪以前的电话服务，你会发现：

> 除或在通信服务背后的最为关键的问题是服务的渗透率。如果其他人都不用这项服务，那么它就没有价值；如果其他所有人都在使用这项服务，那么它就是必需的。

从上述的一个极端转到另一个极端是相当有难度的，这需要公司及个人在一项新的通信技术的经济效益不明显的情况下对其进行投资。如果他们选择的技术得不到普及，那么投资就无法收回。即使其他人采用这项技术，但认为该项技术有较好的经济效益的用户却为数不多。许多人在新技术到来时总是在等待，并观望其他人会怎么做，其目的是希望将自己的经济风险降到最低。对公司而言，从财务上决定电话安装的增长速度更有难度。如果公司安装的电话数太少，那么拨打者可能会收到忙音；如果电话装得太多，那么电话就可能比较空闲，这就意味着公司浪费了资源。

3.7 访问无处不在

为什么美国人最终还是选择了电话服务？如果你是一名历史系的学生，你就应该知道答案：因为美国政府断定普及电话服务对美国具有重要意义，其他大部分国家的政府也做出了同样的决定。美国政府成立了一个正规的垄断组织：美国电话电报公司（American Telephone & Tele-

graph，AT&T），来确保电话服务对于每个家庭及每家公司都可用；同时还规范了电话服务资费，保证普通家庭也能负担得起电话服务费。除此之外，美国政府还要求电话系统既要覆盖城市也要覆盖到乡村的每个角落。更为重要的是，政府鼓励AT&T把所有本地电话服务连接起来，形成一个统一的、庞大的电话网络。

由于仅有一家公司拥有并控制着美国大部分电话网络，所以许多任务都比较容易完成。比如，AT&T能够决定两个城市的电话系统之间互联的技术细节，让一家公司掌控整个电话服务系统使新技术的采用变得容易起来。同时，这也易于实现全局编号系统，以便于一个城市的用户直接拨打另一个城市的用户的电话。

简言之，此项政府行动的结果是电话服务以普通家庭能够负担的价格为广大国民所普遍使用。在后来的几十年时间里，大部分企业和大多数人都用上了电话。当然，没有政府的干预电话服务也可能会得到普及，我们也只能推测那种情况下会发生些什么。最关键的一点不在于政府的干预，而在于电话服务发展成为普遍服务已经成为现实。所有公司都明白普遍电话服务意味着业务流程的转变，当公司及个人都开始安装电话时，每个人都明白电话对自身的重要性，安装电话成为一件急需的事情。最终，电话服务由富人才用得起的奢侈品转变成普通大众都能触及的日常工具。

> 在美国，电话系统成为20世纪选择的通信系统是因为美国政府委托一家公司负责管理普遍电话服务，从而保证其涵盖每个人。

3.8　电话系统与Internet的关系

与电话系统一样，Internet也提供通信服务。当前，Internet正处于由受限访问到普遍服务的艰难过渡期。虽然Internet的接入数一直在迅速增长，但它并没有普及到每个人；尽管美国政府对Internet的发展做出了一定的贡献，但她并没有决定托管普遍服务，因此，与电话系统不同，Internet的成长一直依赖于经济状况，这样导致的结果就是，Internet以一种偶然的方式得到成长。在20世纪90年代中期，一些大公司认为它们可以从Internet连接中获利，于是便开始为自己公司的网站做广告。截至2000年，数百万的家庭已经享受到了低速的Internet访问服务，而各个公司都在升级自己的计算机以及Internet接入技术，目的是有能力处理日益加剧的拥塞问题。到2005年为止，家庭使用高速Internet连接已经变成一件非常平常的事。

在过渡期间，说服那些没有接触过Internet的人，让他们相信Internet会带来激动人心的新体验，这就和在通用电话系统安装之前试图推销电话服务一样困难。而对于那些见过Internet技术的人们来说，他们常常会有礼貌地笑一笑，点点头，但同时心里也在想："它看上去非常不错，但是我又该怎样使用它呢？"这个问题类似于那些没见过电话的人所提出的问题："是的，我知道它如何工作，可是我打电话给谁呢？"

当然，问题的答案是，随着越来越多的人们接入Internet，你自然也会想到用它来联系公司、朋友、银行、学校、政府部门和亲属。下一章将讨论Internet的发展历程，并说明Internet正逐渐普及开来。

第 4 章　过去的模拟技术时代

4.1　引言

Internet 通过数字技术承载各种形式的信息。本章将介绍模拟信息，并说明如何把模拟信号编码成数字形式，并以音频为例进行讨论。

4.2　声音、振动与模拟记录

高速公路的设计者使用简单的机制就可以提醒司机减速。他们在公路上安置了一系列的小凸块，当汽车经过这些小凸块时，车胎会不停地振动，人们把车胎的这种振动感知为声音。

第一台机械式留声机使用与上述方法相同的原理重现声音。一个尖尖的针状物（称作唱针）在一个表面含有许多记录声音的小凸块的圆柱面或圆盘（称作唱片）上移动。当唱针经过小凸块时，它会使一个平整的振动膜发生振动，产生人类能够感知的声音。图 4-1 说明了这个基本原理。

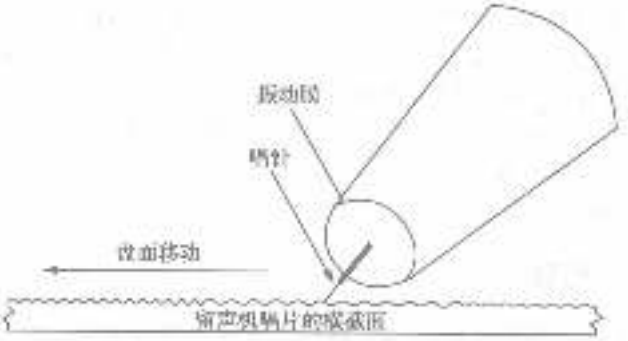


图 4-1　当早期留声机的唱片从唱针下经过时，唱片上的凸块是如何引起振动膜振动的

我们把与留声机类似的设备称作模拟设备（analog device），这是因为这类设备记录和播放的是完全模拟的声音，也就是说，留声机唱片上的小凸块严格地模拟了产生声音的振动。例如，凸块的高度用来控制音量，当声音比较柔美时，凸块几乎是平的；当声音比较洪亮时，凸块就要高些。如果没有一点凸起，那么留声机就不会发出任何声音[1]。

总结如下：

> 模拟设备保持对信息精确的物理模拟。例如，早期留声机唱片上的凸块与人们感知为声音的振动保持一致。

---

[1] 实际上，留声机总会产生一些噪音，因为唱片表面不可能完全平滑，或多或少会存在一些细小的划痕。随着唱片使用次数的增多，划痕也会越来越多。

4.3　模拟电子设备

早期留声机完全是机械式的，而现代的声音播放设备都采用电子方式。例如，调幅广播收音机采用模拟技术，无线电广播站传送声音的模拟信号。当音量比较大时，传送的信号就比较强，否则就比较弱。实际上，理解模拟这一概念的最好方法就是考虑两种物质之间的比例关系：无线电信号的强弱与音量的高低是成正比的。

当我们把收音机调到与发射站相同的频率时，收音机中的电路捕获输入的电波，然后生成与声音的模拟信号一致的电流。当信号与柔美的声音一致时，电流就比较弱，反之则比较强。因此，我们总结如下：

> 如果一种电子设备产生的电流与它的输入信号成一定的比例，那么这类设备就是模拟电子设备。

4.4　许多电子设备都是模拟设备

许多人们熟悉的电子设备都采用模拟技术。例如，调幅或调频广播收音机、立体声音响系统和电视都使用模拟电路。实际上：

> 大多数电子设备都曾经使用模拟技术存储、放大以及发送图像或声音信号。

4.5　早期的模拟通信

模拟通信是早期电话系统的重要组成部分。最初的电话包含两大基本部件：把声音转换成模拟电信号的麦克风与把模拟电信号转换成声音的听筒。每当有人对着麦克风讲话的时候，电信号就沿着电话线将声音的模拟信号传送到另一部电话机，在另一头，听筒把电信号转换回声音。因为电话系统采用的是模拟信号，所以大音量产生的电流也比小音量产生的大。

> 早期电话采用模拟技术将声音从一个地方传到另一个地方；两部电话间的电流强度与音量的大小成正比。

4.6　模拟简单但不精确

模拟设备的原理最容易理解，因为我们做的大多数事情都是模拟的。当一个人用臂力开门时，门的移动模拟了施加在它上面的推力或拉力；人声音大小的变化模拟了施加在声带上的振动力。同样，人们讲话的音调是对拉伸声带的伸展力的模拟。

尽管模拟易于人们理解，但模拟设备还是有缺点的。总的来说，精确模拟所有输入是不可能的。例如，考虑在磁带上记录声响比较大的击鼓声，我们必须在鼓声出现的地方调准磁性。然而，由于磁性调整受可用磁性材料的限制，所以不可能把磁性调整到能记录任意大的声音。此外，磁带在播放时表面会有磨损，因此每次播放音量都会有轻微地减少。

另外，扩音器性能的不完美也是声音不精确的一个原因。模拟电子设备会在不可预知的情况下改变输入信号，比如它可能会扰乱信号，也可能会录制得不准确，或者增加了一些背景噪音。我们把这种改变称作失真（distortion）。比如，当扩音器的音量调到最大时，即使没有输入信号，人们也能听见扩音器产生的背景噪音。总之，结论是：

> 模拟设备易会使输入信号失真并增加噪音。

4.7　在线路上传送模拟信号

每当电流通过一条线路时，部分信号就会丢失。尽管工程师们把这种现象称作信号损失

(signal loss)，但事实上能量并非真正消失，它仅仅是转化成了热能。这种现象对模拟电信号来说是很重要的，因为，当电流通过线路时，部分电能转化成了热能，信号会变得越来越弱。例如，如果电信号内包含声音的模拟信号，当通过一条很长的线路传送后，音量将会比传送前低一些。

对于模拟电话系统，信号损失会引发一个问题，即信号从一部电话传送到另一部电话后将变弱。如果两部电话之间的距离很远，那么信号很有可能弱到根本就无法听到声音。在早期的电话系统中，信号损失问题是非常严重的，以至于电话只能在很小的局部范围内工作。

随着电话服务的普及，电话公司通过在系统中添置放大器解决了信号损失的问题。人们沿着电话线路有规则地安装放大器，以便增强变弱的信号。放大后的信号会有足够的能量传送到下一放大点，最后，信号会顺利地传至目的地。图4-2说明了这一思路。

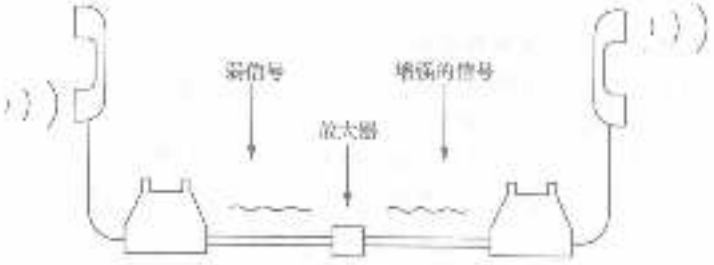


图4-2 如果信号传送距离很长，基于模拟通信技术的电话系统就需要放大信号

当然，模拟电子设备永远都不可能是完美的。两部电话间的线路的每个放大器都会在一定程度上使信号失真，同时会加入一点噪音，而这些噪音将和信号一起被下一个放大器放大。因此，模拟电话系统会包含用于阻塞失真和噪音的过滤器，然而，这样做也意味着同时阻塞了一些正常的声音。另外，在消除噪音的同时，滤波器自身也会使信号一定程度的失真。

4.8 数字音乐

20世纪80年代，音乐产业界开始销售一种称作光盘（compact disc，CD）的音像制品。广告宣称CD音乐的音质优于磁带和黑胶唱片，因为CD是数字（digital）产品。CD刚出现的时候，大多数人都不知道"数字"是什么意思，数字音乐听起来又会怎样，以及为什么它会更好。如果有用户问数字是什么，那么他们将得到如下比较简单的定义：

> 如果一项技术使用数字记录信息，而不是使用像唱片上的凹痕或收音机中的电流一类的物理模拟信号，那么这项技术就称为数字技术。

4.9 数字革命

数字革命并非始于光盘。当前的数字世界之所以能成为现实是由于贝尔实验室的科学家们发明了一种称作晶体管（transistor）的固态开关。过了一段时间，科学家和工程师们又设计出了多种在样品体集成电路[○]（integrated circuit）上集成晶体管的方法。不久之后，这场革命就正式开始了。

集成电路由许多通过线路相连的电子元件组成，它构建在一块边长为零点几英寸的正方形电路板上。经过潜心研究，制造商已经找到了多种减小晶体管尺寸以及制造集成度更高的集成

○ 集成电路的非正式的名称为芯片（chip）。

电路的方法。目前，制造商能生产出包含几亿个晶体管的集成电路。

集成电路的重要性体现在它们的经济价值上，因为集成电路能以低成本进行批量的生产，而对于那些成本高以至于无法使用独立元件的复杂电路就可以采用集成电路方式大规模地生产了。

多数集成电路主要是为计算机设计的，例如，微处理器就是一块集成电路，它是现代计算机的核心部件，其中包括进行加减乘除以及数值比较的电路。除此之外，微处理器还能从计算机内存中提取数据或者将结果存入内存中。

4.10　计算机是数字式的

与前面讨论的模拟设备不同，计算机是数字设备（digital device），之所以称作"数字"是因为：

在计算机内部，所有的数据都用数字表示。

例如，当用户按下键盘上的一个按键后，键盘就会向计算机发送一个数字；当计算机在屏幕上绘制文本或图像时，它也是用数字来实现的。

由于计算机是数字的，所以计算机中的微处理器和其他集成电路都是与数字打交道的，因为计算机电路极其灵活，所以它们可以工作在多种方式下；由于集成电路造价低，所以工程师们已经把它们应用在了许多设备上。例如，微处理器广泛用于掌上计算器、汽车、电视、冰箱、微波炉、移动电话以及办公设备上，它们能控制加热部件、飞机、相机和交通信号灯。

4.11　数字录音

以数字形式录制声音看起来似乎不太可能，毕竟，众所周知，声音是一系列音调和音量不断变化的振动，声音似乎与数字没有一点关系。数字录音能够使用的原因在于计算机内部电路的运行速度比人耳辨别声音快得多，而且计算机还有进行数模转换的机制。

4.12　使用数字重现模拟信号

为了便于理解数字记录，我们考虑一下夏季某一天的温度变化情况。清晨，天气比较凉爽，但是日出之后，气温就迅速升高，到正午前后，气温到达了峰值。临近傍晚，气温开始下降。比方说，你想要在室内重现夏日室外温度的变化，这里假设你有一台加热灯，一个能控制该加热灯所产生的热量的调光开关和一支温度计。

要重现在某一天里温度的变化情况，首先必须把这天的温度变化记录下来。将温度计置于室外，并周期性地记录气温（比如半小时记录一次）。第二天，拿出温度记录表，并把温度计放进加热灯中。通过调光开关，你可以每半小时调高或降低灯的温度，每次调整后的温度值必须与第一天在户外同一时间得到的温度值相同。

计算机电路以数字形式（例如光盘）记录声音时，采用了与上述相同的技术。传统的麦克风产生模拟电信号，该信号沿着电话线传送到数字记录器。在记录器内部，计算机电路周期性地侦测输入信号，并产生一个表示此刻信号强度的数字。计算机电路的运行速度很快，每秒钟能够产生成千上万个这样的数字。这些数字被保存下来，并传送到光盘上。当有人播放这张光盘的时候，光盘播放机内部的计算机就读取这些数字。计算机利用这一系列的数字还原与原始信号差不多的模拟电信号。最后，这些信号经过放大器放大后被送到扬声器。

我们把将模拟信号转换成一系列数字的电子电路称为模/数转换器（Analog-to-Digital converter），简称A/D转换器（A/D converter）。A/D转换器测量模拟电信号，并生成与之匹配的

数值。计算机定时从A/D转换器中读取数字并保存起来。图4-3说明了这一转换原理。

图4-3 执行A/D转换的集成电路。该设备以模拟信号为输入，输出一系列的数值

想要播放光盘，计算机需要一块称为数/模转换器（Digital to Analog converter，D/A converter）的集成电路，如图4-4所示。

图4-4 执行D/A转换的集成电路。该设备以一系列的数值为输入，输出模拟信号

要从数字CD还原声音，计算机先从CD中读取数字，然后把它们送至D/A转换器，再由其生成模拟电信号，之后，该信号经过放大后被送到扬声器播放。由于计算机将数字送到转换器的速度非常快，所以人们耳朵听到的声音是连续的。

4.13 为什么选择数字音乐

从本质上讲，光盘仅仅包含一长串数字，这些数字指定了如何还原出模拟信号。CD如此迷人的原因在于播放的时候不会"磨损"。CD中的数字是经过了编码的，这些数字能被激光器发出的激光感知。因此，在播放CD的时候，没有什么东西触及到CD盘面。而且，生产厂家在制作CD拷贝时，记录盘不会有任何磨损，每张拷贝与源盘一模一样。除此之外，CD中不会掺杂模拟技术中那么多的背景噪音。例如，在两首歌之间，CD仅仅包含一系列的0，即是说，不会产生任何电信号。

4.14 小结

声音源于振动。音频信息最自然的表示法是采用模拟形式，在这种形式里，物理量的变化与声音成一定的比例。比如，留声机唱片上凸块的高度和收音机产生的电流强度都与音量的大小成正比。

音频信息同样能用数字形式表示，即是说，我们用一系列数字表示声音。因为计算机用数字表示一切，所以数字表示信息的方法对计算机而言是最方便的了。

在阅读本书后续章节的时候，读者要牢记两个重要的思想：

- 包括文本、相片、音频和视频在内的任何种类的信息都能用数字形式编码。
- 数字表示法的主要优点在于信息在存储、复制或传输的时候不会出现失真。

第 5 章　数字网络的过去与将来

5.1　引言

本章将讨论数字通信的概念，说明要传输的数字信息如何使用两个基本字符进行编码。

5.2　早先的数字时代

第 4 章曾提到，世界上的大多数电子设备都曾是模拟设备。事实上也的确如此。在光盘出现之前，调幅与调频收音机、立体声音响、电话机以及电视都使用模拟电子电路。令人惊讶的是，最早的电子通信设备并非模拟的。在基于模拟技术的电话普及之前的数十年里，人们使用的是数字通信技术！

近 150 年前，人们就能在几分钟内将一则消息从一个城镇发送到另一个城镇。这项技术称作电报（telegraph）。电报技术出现后很快便得到了普及，以至于电报线路迅速遍及全国。

电报技术的基本原理与墙上控制电灯的开关一样。开关一般位于墙上的一个合适高度，并且它距离电源和电灯一般都比较远。与开关相连的一对电线将电能传送到开关，再把电流送回电灯。当开关处于“断开”状态时，电路断开，没有电流流向电灯；反之，当开关处于“闭合”状态时，电路闭合，电流流向电灯。

基本电报技术也使用带电线的开关进行工作，只不过这个开关控制的距离长得多。即是说，开关位于一个城镇，而其操纵的设备却在另一个城镇。此外，电报不用电灯，用的是一种小型电子设备，该设备接收到电流时会发出滴答声。要通过电报传递消息，发送方必须不停地按动开关，与此同时，在另一座城市里的另一个人在本地设备旁接听耳机发出的滴答声。

对一个未经训练的人来说，电报听起来就像一串人们无法理解的没完没了的滴答声。其中一些滴答声持续时间较短（即，开关被按下的时间很短），而另一些则比较长。有时，在滴答声再次响起之前，会有比较短暂的停顿。训练有素的电报操作员能够从这些滴答声中区分出每个字符。当电报信号到达后，操作员需要尽快把电报转换成纸写的消息。

5.3　电报是数字的

电报机是数字设备，因为它不是发送输入信息的准确的模拟信号，而是通过滴答声发送消息中的每一个字符。尽管电报的滴答声似乎与数字无关，也与数学家无关，但是这些滴答声代表的是一组数字系统的所有数字。事实上，我们现在可以更加准确地定义数字式（digital）：使用有限离散值集合的任何设备是数字式的。总结如下：

> 电报技术是数字技术，因为它传送离散的滴答声，而不是连续变化的信号。

5.4　莫尔斯码

Samuel Morse 发明了一种后来在电报操作员中普及的编码，莫尔斯码。这是一种用一系列滴答声和停顿表示字母、单词和简单的方法。例如，莫尔斯码用一短后接一长的滴答声表示字母 A。

Morse在分配码值时，用短音序列表示常见字母，这样设计的好处在于其发送消息的速度比没有经过精心设计的电码更快。例如，在英文文章中，字母E出现的频率高于其他字母，Morse用一个短促的滴答声来表示E。

短滴答声称作点（dot），而长滴答声称作划（dash）。在传送的过程中，在表示一个字母的所有点和划之后有一个短暂的停顿，而在每个单词之后有一个较长的停顿。经过训练的操作员根据这些停顿来识别各个字母和单词在哪里结束。

5.5 莫尔斯码中的字母与数字

除了对所有字母进行编码之外，Morse还对0到9这十个数字以及一些标点符号进行编码，如图5-1所示。

A	·–	X	–··–
B	–···	Y	–·––
C	–·–·	Z	––··
D	–··	0	–––––
E	·	1	·–––––
F	··–·	2	··–––––
G	––·	3	···–––––
H	····	4	····–––––
I	··	5	·····
J	·–––	6	–·····
K	–·–	7	––·····
L	·–··	8	–––·····
M	––	9	––––·····
N	–·	,	––––––·
O	–––	.	·–·–·–·
P	·––·	?	··––···
Q	––·–	!	·–·–·––
R	·–·	:	–––···
S	···	;	––––––·
T	–	-	·····–
U	··–	)	·––––·
V	···–	)or(	–·––––·
W	·––		

图5-1 莫尔斯码示例图。莫尔斯码使用惟一的点、划序列表示每个字母、数字与标点符号

Morse并没有给对现实中存在的所有符号分配编码，比如，美元符号和百分号就没有对应的编码，即使这些字符在书面文字中的的确确有可能出现。

5.6 用户不会接触莫尔斯码

虽然通过电报发送的所有消息都采用莫尔斯码进行编码，但只有电报操作员需要熟悉莫尔斯码，比如，有人想要发送一份电报，他把要发送的消息写在纸上，然后交给电报操作员，剩下的工作就由操作员完成了。消息本身可以用任何语言书写，实际上，由于电报供应商按发送消息的长度收费，所以人们常常采用单词的缩写形式。

训练有素的电报操作员能够在莫尔斯码和消息间迅速转译，通过电报系统传送消息至少需要两个操作员。在发送方，操作员把纸上的电报原文译成莫尔斯码并发送出去；在接收方，操作员收听莫尔斯码并将其转译成的文字记录下来。最后，把接收到的消息投递给最终接收方。

电报技术中有三种思想与Internet相关，如下：

- 仅用两个基本编码值（点和划）就可以对所有的字母数字进行编码。
- 用于发送信息的代码定义了一套字符编码表，虽然该表没有包含所有可能的字符，但是它仍然十分有用。
- 使用电报服务的客户完全不会接到，也不用懂得底层的编码策略。

5.7　事实上的即时通信

发明电报后，人们认为它似乎拥有其所魔力，在这之前，把消息送到遥远的地方需要专门的送信人，而且这些送信人常常是骑马送信的，突然间，整个世界发生了转变，人们有可能在事件发生后的较短时间内了解到事件的相关信息，比如，通过电报，远离金融市场的股民可以了解当前股价信息，也能够发送购买或抛售股票的指令，远离计票点的选民能够即时地知道选举的结果，旅行者也能跟家人或朋友保持联系。

5.8　速度是相对的

与专门的送信人相比，电报要快得多，它改变了整个世界，但是人们还是觉得电报通信相对比较慢。想像一下用电报代替电话来跟朋友进行通信，写完消息后，你必须把它交给电报操作员，然后等待操作员把这些消息翻译成莫尔斯码。要知道，最熟练的操作员每分钟最快发送十几个单词，而且，发送和接收操作员都必须同等熟练方能保证传送的成功。如果接收方错过一个字符或单词，那么他就只能要求发送方重发一遍。因此，我们容易得出这样一个结论，通过电报进行对话既不方便也比较慢。

电话如此吸引人的原因是显而易见的，在电话一端的人可以直接与另一端通话，而不用把消息写下来，然后交给操作员。电话系统即时地把讲话者的声音传送到另一端，并且它还能传输不易书面表达的内容，如情绪、闻其声，接听者很容易区分愤怒与诙谐、沉默与兴奋。

电话通信很快便流行起来，许多从事通信系统研究工作的工程师摒弃了较慢的数字电报技术，转而研究快得多的电话模拟技术。

5.9　数字电话

尽管声音通信本质上讲是模拟的，但现代电话系统都采用数字编码传送声音。系统在发送方把模拟的声音信号转换成一系列数字信息，这点类似于光盘。计算机通过电话网络传送这些数字信息，当到达目的端后，系统再把数字信息还原成模拟音频信号。

使用数字技术传送声音给电话公司带来了重大利益。为了理解这一点，我们看一下电话呼叫过程。最初，长途电话呼叫需要一座城市的接线员去联系另一座城市的接线员，他们通过协商将两座城市的电话线连接在一起，以此实现通话。后来，AT&T采用拨号技术，取代了人工接线的方式，实现了线路之间的自动连接。要建立呼叫，用户仅仅需要拨打目的方的电话号码。用户拨号时，电话机会发出数字脉冲，拨号系统利用这些数字脉冲将该电话机连接至目的电话机。

在拨号机制刚刚建立起来的时候，电话公司使用两组电话线，一组传送用于拨号的数字信号，另一组则传送声音的模拟信号。过了一段时间之后，工程师们将语音系统和拨号系统合并在一起，最终，现代电话系统几乎完全是数字式的。

5.10　与 Internet 的关系

与早期的电报类似，Internet也采用数字通信技术。由于计算机以数字形式存储各种信息，所以数字通信技术能够很好地在计算机网络中运行。当信息从一台计算机传送到另一台计算机时，数字机制节省了大量的时间和精力。

5.11　Internet 中数据的二进制编码

Internet和电报还有另一个相似的地方：它们都仅仅使用两个值来编码所有的数据信息。莫

尔斯码中使用的两个值通常称作点和划，而 Internet 中用到的两个值分别是 0 和 1，即二进制系统的两个"数字"。与大多数计算机系统一样，Internet 中的这两个数字也称作"比特"(bit)[⊖]。下一章将介绍用于 Internet 上的与莫尔斯码作用相同的编码技术，并解释如何用比特序列表示单个字母。

5.12 为何仅使用两个字符

数字编码技术采用两个符号不仅仅局限于 Internet，所有的电子数字设备都用比特序列编码数据。为了理解这一点，我们考虑在薄雾中通信的情况。假设你有一只手电筒，其他人很容易就能分辨手电筒是"照亮"还是"熄灭"，然而要区分手电筒是"熄灭"、"暗淡"、"适中"还是"明亮"就困难得多了。同样，相对于区分多个信号状态，电路分辨"打开"与"关闭"显然容易得多。

即使 Internet 以比特编码数据，这些技术细节对用户来说都是不可见的。与用电报发送消息或用光盘播放音乐的人们一样，Internet 用户从来也见不到底层的二进制编码机制。

> 虽然 Internet 使用二进制编码机制传输数据，但用户往往并不知道这种编码机制。

5.13 小结

Internet 和其祖先——电报有三方面的相似之处。第一，Internet 也提供数字通信服务，允许人们在计算机间传送一些数字。存储在计算机中的数字可以用来编码几乎所有信息，包括文档中的字母、声音、图像等。第二，Internet 跟电报一样在底层使用两个值来编码所有数据，只不过 Internet 用的是 0 和 1 这两个二进制数字。第三，Internet 对用户隐藏了数据编码细节。

⊖ 术语 bit（比特）是 binary digit 的缩写。

第6章 基本通信

6.1 引言

计算机网络（computer network）将若干台计算机互联起来，以便它们之间相互交换数据。虽然现代计算机网络是硬件和软件的复杂组合体，但早期的计算机网络要简单得多。

本章将概述基本通信技术的发展，并介绍网络演变的过程。本章还将介绍一些术语，并阐述调制解调器的工作原理。在本章定义的概念对以后的章节很重要。

6.2 利用电进行通信

自从发现电之后，发明家、科学家以及工程师们就一直致力于研究通过电信号进行通信的各种方法。利用发现的这些原理，人们构建了快速、可靠的通信系统。人们对数字通信的认识大致可以分成三个阶段。第一个阶段重点研究线路中信号的特性；第二个阶段着重研究如何使用电信号发送比特以及如何用比特表示字符；第三个阶段的重点是研究如何检测和校正传输过程中出现的错误。

6.3 线路上的信号

研究人员首先研究电信号是如何在导线上传送的。例如，他们发现电信号从金属线终端反射的方式与光线从镜面反射的方式相同（这是许多现代网络需要在每条导线终端配置终结器的原因）。他们还发现电信号在传送的过程中能量在不断减少（这是现代网络限制互联导线长度和使用电子设备放大信号的原因）。此外，他们还发现导线内的电信号会产生电磁辐射，它会干扰附近导线内的信号（这是高速网络和有线电视采用把导线封装在金属护罩里的专用电缆的原因）。

6.4 信息编码技术

研究人员知道如何在导线上发送电信号之后，他们便开始着手研究用电信号编码信息的各种方法，其中大量的研究工作集中在寻找语音编码方案上，这类方案能有效地在电话线上传送语音信息，而且能将传输过程中的信号失真减至最小。值得注意的是，多数编码技术普遍适用于各种通信。

研究人员发现一种能很好地发送语音信息的技术，人们把这种技术称作调制（modulation）。要使用调制技术，发送方必须配置一台调制器（modulator）。调制器首先产生一条上下规则振荡的基本电信号，这种基本振荡信号称为载波（carrier）。调制器用另一个输入信号（如一个人对着话筒讲话发出的声音）稍微地改变载波。在线路的接收方，解调器（demodulator）反向执行功能，称作解调（demodulation）。解调器首先要调整好，等待接收载波信号，通过测量输入载波与标准载波的偏离量，接收器就能够恢复输入信号（人的声音）。在确定人的声音的强度范围后，科学家们找到了足以传送音频信息的载波信号。

调制解调器

调制的原理在现代通信系统中仍然有用。接入 Internet 的两台计算机可以通过一条长电缆进

行通信，在通信前，一端需要一台调制器，而另一端需要一台解调器。用户可以在建筑物内安装线缆，也可以从电话公司租用建筑物间或者穿越城市甚至全国的线路。租用的线路不会直气地接入电话系统，无论谁租用了专线都只能用于私人通信。当用户采用了一套通信线路后，他必须在线路两端配置一种设备，这种设备称为调制解调器（modulator/demodulator，modem）。实际上，一条专用线路总是包含两组独立的线路，每组线路用于一个方向的数据传输。调制解调器中既包含发送信息的调制器，也包含接收信息的解调器。图6-1说明了这一概念。

图6-1 调制解调器采用调制技术在传输线路上传送数据的示意图。计算机在与调制解调器交互的过程中发送和接收数字数据；调制解调器对要传输的数据进行编码

还有一种调制解调器称作拨号调制解调器，它可以在普通拨号电话线路中传送数据。拨号调制解调器包含了附加电路，可以拨出号码和应答呼入。用户必须指示一端的调制解调器拨号，而另一端的调制解调器应答呼入。一旦连接建立，两端的调制解调器就可以开始传输数据了。

拨号调制解调器在编码数据时使用人耳可感知的音频作为载波，这种载波就是调制解调器工作时人们在电话听筒里听到的高音频声音。每当计算机发送数据时，发送方调制解调器通过改变音频强度来调制载波音调。接收方调制解调器监听载波信号，在载波发生改变时便对数据进行解码。

6.5 调制解调器支持双向通信

调制解调器允许数据在它们之间双向传输。每台调制解调器都包含调制部件和解调部件。此外，两端的调制解调器可以使用两种不同的载波信号，或者是轮流发送数据。无论如何，数据都在两个方向上同时传送。

总结如下：

> 调制解调器是一种用于通过拨号电话连接的通信并通过线路进行长途通信的设备。调制解调器包含用于发送数据的调制器和用于接收数据的解调器，所以它支持双向通信。

6.6 数字信息的字符编码

研究人员在研究声音传送的同时也考虑到了数字信息的传输。他们找到了多种用电信号编码基本比特位的方法，例如，用正电平表示1，负电平表示0。除此之外，他们设计了用比特序列来编码各个字母和数字的方法。

尽管现代计算机网络中采用的字符编码仅仅用到了两个基本数值，但是这些编码与莫尔斯码还是有不同之处的，因为每个字符都由相同比特位数的编码表示。例如，莫尔斯码用单一点表示字母E，而用3个点表示字母S。相比之下，许多现代字符编码都用0、1来表示每个字母，所有字符都具有相同比特位数，这一点使字符处理的速度更快，硬件也更加便宜。这也简化了字符存储，因为每个字符占据的空间大小都是一致的。

美国信息交换标准代码（ASCII，American Standard Code for Information Interchange）是目

能最流行的字符编码之一，ASCII 码广泛应用于计算机与网络领域。ASCII 码定义英文中使用的大部分字符，包括大小写字母、数字、标点符号以及诸如大于、小于、等于、加和减等数学符号的各类符号。

由于使用计算机或网络的大部分人从来都不曾见过 ASCII 码，所以 ASCII 码的编码细节并不重要。但是图 6-2 的例子有助于人们弄清 ASCII 码的编码原理。ASCII 码用 7 位二进制序列 1000101 表示字母 E，用 1010011 表示字母 S，0101100 表示逗号。

总结如下：

> 许多网络在发送数字形式的文本信息时使用 ASCII 编码。ASCII 码给每个字母和数字分配一个 7 位二进制序列。因为 ASCII 码的细节是隐藏在内部的，所以大部分用户并不会接触到它。

A 1000001	S 1010011	a 1100001	s 1110011
B 1000010	T 1010100	b 1100010	t 1110100
C 1000011	U 1010101	c 1100011	u 1110101
D 1000100	V 1010110	d 1100100	v 1110110
E 1000101	W 1010111	e 1100101	w 1110111
F 1000110	X 1011000	f 1100110	x 1111000
G 1000111	Y 1011001	g 1100111	y 1111001
H 1001000	Z 1011010	h 1101000	z 1111010
I 1001001	0 0110000	i 1101001	. 0101110
J 1001010	1 0110001	j 1101010	, 0101100
K 1001011	2 0110010	k 1101011	? 0111111
L 1001100	3 0110011	l 1101100	) 0101001
M 1001101	4 0110100	m 1101101	(1111011
N 1001110	5 0110101	n 1101110	/ 0101111
O 1001111	6 0110110	o 1101111	& 0100110
P 1010000	7 0110111	p 1110000	+ 0101011
Q 1010001	8 0111000	q 1110001	- 0101101
R 1010010	9 0111001	r 1110010	= 0111101

图 6-2　ASCII 码示例图。大小写字母各有一个编码，用户不会见到这些编码

6.7　差错检测

多数数字通信领域的前期工作都集中在差错检测和校正上。研究人员对发送电信号时出现的错误进行仔细研究，最终找到了各种检测错误的方法。比如，人们发现，诸如闪电之类的自然现象会使线路上产生随机的电信号，这些信号会和携带信息的信号混在一起。人们还发现线路上的电信号在通过强磁场的时候（比如，靠近家电的电动马达时）也会出现失真。

当电磁干扰破坏了线路上的信号时，数据会被破坏或丢失。例如，如果用电压表示一位数据，那么一道闪电即便不是直接击中导线，而是击在线路附近，也会引起电压的变化。我们要记住这一点：

> 当使用电信号传送数字信息时，电磁干扰会引发一位或更多位的数值发生改变。

为了保护信息免受随机电子噪音的干扰，研究人员设计了一些检测和校正错误的机制。例如，他们发现，如果在字符编码末尾添加一个附加位，那么就能检测出小错误。在这种机制中，如果字符编码含有奇数个 1，那么附加位置为 1；如果含有偶数个 1，那么该位置为 0。从本质上讲，发送器设置这个附加位的目的就是保证每个字符编码都包含偶数个 1。

这里的附加位称作奇偶校验位（parity bit）。例如，字符 E 的 7 位字符编码是 1000101，其中包含奇数个 1，因此其奇偶校验位为 1。然而，字符 S 的奇偶校验位置 0，因为它的编码 1010011

含有偶数个1。

要使奇偶校验发挥作用，接收器必须对接收到的每个字符检测奇偶性，接收器检测接收到的所有比特位，包括奇偶校验位。如果接收器发现字符编码（包括奇偶校验位）中有奇数个1出现，那么它就会报告出错；否则报告接收到的字符是正确的。如果电子噪音改变了某个正在传输的字符的一个比特位，那么接收器就会发现出现了奇数个1，从而认定这个字符被破坏，然后把它丢弃掉。

虽然奇偶校验有助于查出小错误，但是它并不能保证检测出所有的错误。要解释这一点，这里考虑一个在导线上传送的字符，随机的电子干扰能够引起某些位由1变成0，或由0变成1。如果碰巧电子干扰把这个字符中偶数个1个变成0，那么该字符的奇偶性看起来仍然是正确的。与之类似，如果噪音把偶数个0变成了1，那么奇偶性看上去也是正确的。一种极端的情况就是，如果一个强磁场把一个字符的所有位都变成了0，那么这个字符的奇偶性就一定是正确的了。现总结如下：

> 传输前，在每个字符编码后添加一个奇偶校验位有助于硬件检测字符在传送过程中出现的错误。但是，仅有奇偶校验并不能检测出所有可能的错误，我们需要功能更强的差错检测技术。

实际上，Internet中采用的差错检测技术能够检测到仅用奇偶校验无法检测到的错误。在传输过程中仅有少数情况会出现比特序列遭到破坏，接收器会报告错误，并交由通信软件处理这个错误。

Internet软件是如何检测到破坏了的比特位的呢？软件采用被称作校验和（checksum）的差错检测技术。Internet软件发送的不是一个附加的奇偶校验位，而是将发送消息的数值之和。比如，如果一则消息由数值1、3、5组成，那么该软件将发送4个数值，而校验和是紧跟在最后一个数值之后的：

1 3 5 9

当消息到达接收方时，接收方软件会把除最后一个数值之外的所有数值加起来，然后检查这个值是否与最后一个数值相等。如果这些数值中的任何一个在传送的过程出现了差错，那么计算出来的结果就不会相等。

6.8 小结

研究人员研究了电信号在导线上的特性，懂得了如何使用电信号编码诸如声音信号之类的信息。除此之外，研究人员还找到了利用电信号按比特序列编码数字信息的方法。他们设计了一种给每个字符分配一个惟一比特串的代码。特别是，研究人员设计了在许多计算机以及计算机网络上广泛使用的ASCII码。

研究人员们设计了一种极其重要的传输方案，该方案使用调制技术来编码待传数据。调制首先需要产生载波，即一个有规则的振荡信号（如人们可以听见的声音）。然后用待发送的信息轻微地调制载波。接收器根据输入信号与标准载波的偏移量来还原信息。

提供调制和解调技术的设备称作调制解调器，调制解调器用于在导线上长途传送信息或通过拨号连接发送信息。目前，调制解调器在几乎所有的Internet传输线路上使用。

研究人员还研究了传输过程中可能出现的差错，并找到了有效的差错检测方法，比如奇偶校验策略。当电子干扰对传输过程中的数据产生破坏时，奇偶校验策略可以用于检测差错。虽然奇偶校验有助于差错检测，但是它无法解决所有的问题。Internet采用了更加强大的差错检测技术，比如校验和。

第7章　局域网时代的到来

7.1　引言

人们需要更好的电话通信服务，在此需求的激发下，许多早期的通信研究工作都集中在如何使通信服务覆盖大面积的地理范围上。实际上，研究人员现在仍然在研究远距离通信问题，寻找改进传输性能的方法。20世纪60年代末到70年代初，一种新型网络技术的出现给普通百姓的生活带来了更为直接的影响。本章将介绍这类新型网络技术，并说明这种技术是如何改变计算机行业的。

7.2　动机

这种新技术称作局域网（Local Area Network，LAN）技术，它的兴起主要源于两种趋势。第一，计算机变得更小更便宜，人们开始用小且便宜的小型机（minicomputer）取代体积庞大、造价高达数百万美元的大型机。第二，人们开始意识到计算机可以帮助人们完成日常工作中的大量任务。

低成本小型机改变了人们对日常事务的处理方式。在一台计算机造价超过一百万美元的时候，大部分机构只会配备一台计算机。然而，在计算机价格降到足够低之后，每个机构都能从拥有的几台计算机中获益。显然，拥有许多计算机的机构需要在计算机间相互传送数据，以便实现信息共享。

7.3　可交换介质

起初，在计算机间传送数据需要使用移动介质存储设备，这些设备通常指的是磁带或磁盘。尽管早期磁盘体积很大，但它的工作原理与现代DVD相似。操作员首先把一张空盘放入与一台计算机相连的磁盘驱动器里，然后操作计算机把数据写入磁盘。之后，操作员将这张盘移到另一台计算机的磁盘驱动器里，再命令第二台计算机读取数据。

7.4　计算机由集成电路组成

要了解计算机的组成，必须首先懂得计算机是如何制造的。在计算机内部，电子部件集成在一块薄薄的、平滑的矩形电路板（circuit board）上。每块电路板上既包含各个电子部件，也包含电子部件之间的线路。

所有计算机内部的电路板不尽相同，其有代表性的是，每台计算机上都有一块被称作主板或母板（motherboard）的电路板。当有人购买计算机时，他必须在各种各样的电路板中做出选择，选择结果直接决定着这台计算机将包含哪些附加电路板，人们把这些附加电路板称作子板（daughter board）。例如，有的人可能想要为自己的计算机购买一台打印机和一台摄像机，而有的人则选择的是一台CD播放机和两块硬盘。最后，销售商通过在计算机上安装相应的电路板来满足用户的要求。

7.5 将电路板插入计算机

为了使安装子板比较容易，计算机生产厂家在计算机内部配备了一套插槽。主板上的线路通过这些插槽与计算机的其他部分连接起来，比如，每个插槽中都有一些线路将电能传送给电路板，其他线路用于传送数据。图7-1给出了计算机内部插槽的示意图。

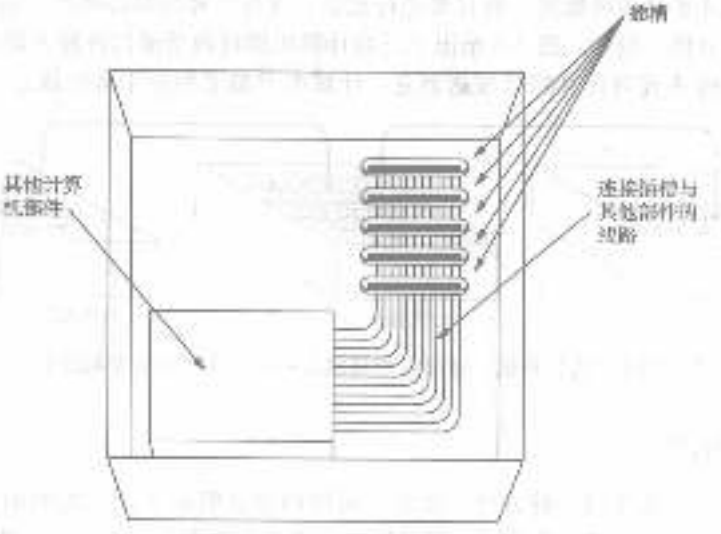


图7-1 移去面板之后的计算机内部组件示意图，子板能插进每个插槽，线路连接插槽与其他部件

每个子板都有一个与主板上的插槽完全匹配的插头。此外，控制外部设备（如打印机）的子板都有一条附加电缆连接到外部设备。组装计算机的意思就是把一组子板插到计算机的插槽上。插上子板并不困难，许多拥有个人计算机的人都通过添加或更换计算机中的子板来升级自己的计算机。

7.6 连接两台计算机

工程师们制造的第一块在两台计算机间用电信号传送数据的硬件是由两块用电缆连接的子板组成，如图7-2所示。

一旦将通信电路板插入计算机，并通过电缆连接到计算机后，这两台计算机就可以传输数据了。子板的工作原理类似于输入/输出（I/O）设备（如磁盘），要进行通信，一台计算机先把数据写到它的子板上，这就好像往磁盘上写数据一样。接着，两块子板上的电路协作地在两台计算机间传送这些数据。最后，另一台计算机从自己的子板中读取数据，就好像从磁盘上读取一样。

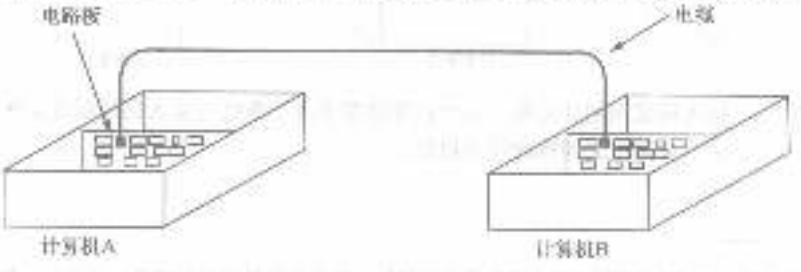


图7-2 早期计算机通信系统示意图，该系统由两块分别插在两台计算机插槽内的子板组成

两台计算机间直接相连的主要优点是数据传输速度快。因为计算机能够提供访问计算机内存的最快路径，所以工程师们就能够制造出在两台计算机间高速传输数据的电路。

除了上述优点，直接相连的主要缺点是不够方便而且费用高。由于必须在每对互联的计算机内安装新的、价格昂贵的电路板，所以这种方式的费用很高。此外，在一组互联的计算机中添加一台计算机是很麻烦，所以这种方式也不方便。况且，如果两台计算机不一样，那么找到一对连接它们的子板几乎是不可能的。而且要进行通信，所有计算机都必须处于运行状态，计算机间直接互联也很不方便。例如，图 7-3 给出了三台计算机通过两条通信链路互联的示意图。从图中可以看出，计算机 A 要向计算机 C 发送消息，计算机 B 就必须处于运行状态。

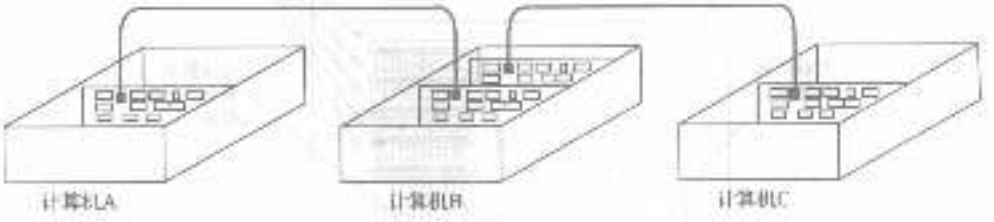


图 7-3　几对对子板连接三台计算机。通信组中每加入一台计算机就需要额外的一对子板和一条线缆

7.7　局域网技术

局域网（LAN）技术以一种方便、便宜、可靠的方式解决了计算机间的通信问题。局域网技术摒弃了计算机间直接相连的方式，而采用专门硬件互联多台计算机。“网络”是独立于计算机本身而存在的，一台接入网络的计算机失效不会影响其他计算机间的通信。

局域网技术用于小范围内的通信。一般来讲，人们可以认为局域网是“在一座建筑物内”。由于局域网仅仅需要覆盖很小的范围，所以电信号在传送到局域网末端的时候仍然有可能保留足够的强度。因此，局域网不需要放大器增强信号强度。

许多局域网系统都使用电缆将每台计算机连接到一台小型电子设备上，该设备称作集线器（hub）或交换机（switch）[3]，它形成了整个局域网的中心。图 7-4 说明了多台计算机是如何连接到局域网中的。

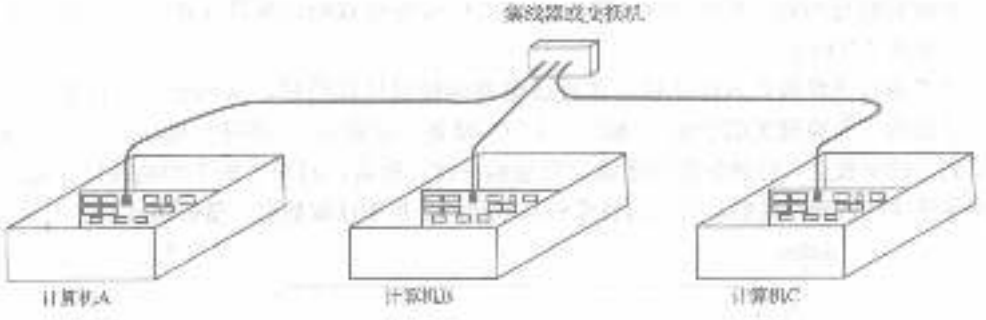


图 7-4　接入局域网的计算机。每台计算机都通过一条电缆接入集线器或交换机，之后计算机之间就能直接通信了

[3] 从概念上讲，集线器和交换机的工作原理是一样的，但是交换机的费用较高，并且在一些情况下，它能够更快的发送数据。

在一栋办公楼里，将计算机接入局域网的所有电缆一般都铺设在走廊大花板上。每条线缆都会插入集线器或交换机。

局域网的具有很高可靠性，要使整个局域网系统更加健壮，正常的运作就不能依赖于所有计算机和所有硬件线路都要工作正常。即使一台计算机中的子板失效或者一台计算机与局域网的连接断开，其他计算机仍然能够继续通信。

总结如下：

> 如果一种计算机网络技术提供小范围内计算机之间互联的方法，那么这种技术就属于局域网技术。局域网具有造价低、可靠性高以及安装和管理方便的优点。

7.8 将计算机接入局域网

计算机在接入局域网前需要网络接口硬件。从计算机的角度来看，局域网接口的工作方式与I/O设备接口相似。从物理上讲，网络接口硬件由两部分组成：一块通常被称作网络接口卡（Network Interface Card，NIC）的子板和一根将NIC接入局域网的电缆。NIC也需要插入计算机的插槽中。

计算机的处理器通过NIC访问局域网，它可以请求NIC发送一则消息给局域网上的另一台计算机或者从NIC中读入输入的消息。

> 计算机需要附加的硬件来建立与局域网的连接，这些硬件由插在计算机主板上的网络接口卡（NIC）和连接NIC和局域网的电缆组成。成功连接之后，计算机就可以通过NIC发送和接收数据了。

局域网是独立于与其相连的计算机的，例如，数据的格式与数据在局域网上传送的速率并不依赖于这些计算机。每台计算机中的网络接口卡负责把数据转换成局域网要求的格式，并协调计算机与网络之间的速率差异。要达到上述要求，NIC中通常包含一个微处理器以及容量很小的高速随机访问存储器。NIC利用高速存储器缓存输入和输出的每条消息，以便这些消息能够以网络所用的速率发送给局域网或者以计算机所用的速率传送给计算机。

举一个例子，当计算机A发送一则消息给计算机B时，它并不是直接把消息从内存传送到局域网，而是把消息发送给它的网络接口卡。然后，A的NIC通过局域网并以局域网要求的速率把消息发送给B的NIC。在B的NIC接收到这则消息之后，它把消息复制到B的内存中。因为仅仅是网络接口卡用来发送和接收数据，所以这两台计算机既不需要以相同的传送速率工作，也不需要跟局域网的传送速率一样，这种重要的想法即：

> 由于接入局域网的每台计算机都包含有将计算机和局域网隔开的网络接口硬件，所以局域网的速率并不依赖于与其相连的计算机的速率。因此，各种不同的计算机都能通过局域网进行通信。

7.9 局域网技术的重要性

局域网技术改变了人们使用计算机网络的方式。在局域网技术出现之前，人们认为计算机通信是一种在较大地理范围内使用的通信方法。但是，局域网技术出现之后，人们开始利用网络连接一个房间或一栋建筑物里的机器。

局域网技术带来的意义最具深远的改变就是实现了资源共享。该技术出现之前，计算机之间都是自我封闭的。每台计算机都有一组I/O设备，比如打印机和磁盘；每台计算机都有一份用户能够访问的软件。一旦使用了局域网技术，一组计算机之间就可以相互共享诸如打印机这类

的资源了。

资源共享在很大程度上改变了计算机产业，由于网络连接比一组I/O设备要便宜得多，所以将许多计算机接入网络，并利用网络来共享I/O设备，这显然是明智之举。总之：

> 局域网改变了整个计算机产业，因为它使通过便宜的计算机共享诸如打印机和磁盘之类的资源成为可能。

7.10　局域网与Internet的关系

当Internet工程启动的时候，局域网技术才刚刚萌芽。Xerox公司将一种新型局域网技术的样板提供给了几所大学，该样板称作以太网（Ethernet），由Xerox公司的一个研究室开发，此项技术注定成为领头的局域网技术。

Internet的研究人员构想了一幅未来的蓝图，他们认为局域网技术会变得很便宜，而且会得到广泛的应用。例如，人们设想各个团体和组织都会用一个或多个局域网连接它的所有计算机。研究人员根据这份蓝图设计了Internet，后来的事实证明了这些设想的正确性。

第二部分　Internet 简史

Internet 发展简史

Internet 如何从起步时的
微不足道发展成
世界上最大的网络

第 8 章　早期的 Internet

8.1　许多独立的网络

20 世纪 70 年代后期，计算机网络开始兴起。几家计算机生产厂家推出了具有足够计算能力来处理多用户的小型计算机。由于这类计算机价格并不昂贵，所以大型团体的各个部门都能负担得起这笔费用。

为了实现小型机之间的互连及这些计算机间数据的高速传输，许多团体开始搭建局域网。因为局域网技术既便宜又易安装，所以单个部门无需咨询中央管理部门就可以购买、安装和运行一个局域网。

8.2　数量激增的局域网

允许团体中各个部门自行搭建并运行自己的计算机网络既有优点也有缺点。当到有自主权后，单一部门就能够自行选择适合自己需要的网络技术。他们能对安装和允许局域网的资金做出预算，也能决定谁有访问权限，制定网络使用的规则。

当然，自治也带来严重的缺点，其中最为重要的是，由于不是所有的厂商都提供各种类型网络所需的全部硬件，所以自治会鼓励各种局域网技术数量的剧增。

8.3　有关局域网的几点事实

要理解局域网为什么会激增以及怎样激增，我们必须了解下述三个事实。

(1) 工程师们已经设计了多种局域网技术。

局域网硬件可以按照速度、可靠性、易安装、容量和成本 5 个因素的某种组合来设计。工程师按不同的组合已经设计完成了几种网络硬件。

(2) 局域网的性能决定成本，高性能局域网成本较高。

人们大概都知道，传送数据较快的硬件的价格也较贵。因此，一个团体可能通过选择性能较低的局域网来节省费用。的确，仅以很少的费用搭建局域网是可能的。

(3) 特定的局域网技术可能只适合特定的计算机。

上文提到，把一台计算机接入局域网需要网络接口硬件。计算机生产厂家可能无法提供所有类型局域网的接口硬件，尤其是，当计算机生产厂家开发一种新的局域网技术时，它通常只销售与自家计算机匹配的接口。即使局域网技术由不生产计算机的公司开发，该公司通常也只销售时下最流行的计算机所使用的总线接口硬件。

上述最后一条表明，一个团体在决定搭建局域网时，其选择范围可能会因产品的可用性而受限。如果该团体拥有一批计算机，那么他们必须验证是否每一台计算机的接口硬件都可用。第二条表明局域网技术的选择可能取决于经济状况。如果一个组织中的两个部门的可用资金不等，那么它们可能选择不同的局域网技术。上述第一条表明有许多可供选择的局域网技术。

由上述事实得出的结论很简单：由于一个团体中的各个部门可以选择满足自身需要的技术，所以大部分大型团体都同时使用多种技术，其中一些技术是为了高速，另一些是为了易于安装

和维护，其他则是为了最大限度的节约资金。总结如下：

> 选择哪种局域网技术依据的是速度、易用性和特定计算机接口的可用性。一些大型团体同时使用多种局域网技术。

8.4 局域网是不兼容的

局域网技术的缺点可以简洁地概括为一句话：

> 各种局域网技术之间是完全不兼容的。

这句话的意思就是多个局域网不能接插在一起，例如，假设一个团体部署了两个局域网，一个位于运输部门，一个位于财务部门。再假设有人需要把存放在运输部门计算机中的数据传送到财务部门的计算机里去，仅仅连接两个局域网的电缆并不能解决这个问题。

从技术上讲，多个局域网无法相连的原因有如下几个：第一，每种局域网技术都在有限的范围内有效，每种技术都指定了最大电缆长度（例如，一种流行技术指定其电缆长度必须不超过500米）。扩大局域网的覆盖范围有可能导致网络功能失常。第二，每种局域网技术都有对应的电气规格，比如，电压和频率，因此，不同的局域网技术可能在电气规格上相互不兼容。第三，每种技术都用自己的信息编码方法，比如，调制的方式。一个局域网系统使用的编码方式对于其他局域网系统来说一般都是没有任何意义的。

8.5 广域网技术

20世纪六七十年代，除了局域网技术，另一种形式的计算机网络也出现了。科学家和工程师们设计出了各种构建将分布在广大地理范围内的计算机连接在一起的网络的方法。这种远距离通信技术称作广域网（Wide Area Network，WAN）或称作远程网（long-haul network），广域网跟早期的远程通信系统使用相同的基本机制。例如，广域网利用调制解调器在长距离传输线路上发送信号。但是，广域网不只是通过一条传输信道把两台计算机连接起来，而且还利用计算机将一组传输线路整合成了一个有序的系统。为了达到这个目的，广域网中的每个站点都包含了一个小型的专用计算机，该计算机与传输线路相连，并使网络运作独立于使用网络的计算机。这个专用计算机接收来自另一个站点的输入信息，然后把信息发送给位于本地的一台目的计算机。同时，它也接收来自本地任何一台计算机的数据，然后通过传输线路将这些数据传送给目的计算机。总结如下：

> 广域网技术（WAN）不同于一组分立的传输线路，因为广域网在每个与传输线路连接的站点安置了一台附加的、具有特定目的计算机，该计算机使网络运作独立于使用广域网的计算机。

为了理解广域网的用途，想像一下，有一家公司在纽约、芝加哥、洛杉矶和奥斯丁4座城市拥有办事处。该公司可以通过安装一个广域网来连接各个办事处的计算机。这个广域网在物理上可能由三条租用的传输线路组成：一条从芝加哥到纽约，一条从芝加哥到洛杉矶，第三条从芝加哥到奥斯丁。此外，在每个连接点安置一台小型专用计算机。从概念上讲，广域网的功能与局域网很相似，只不过运行速度慢一些。广域网允许所有接入的计算机相互通信。

广域网理解传输线路互连计算机的方式，能够自动处理数据传送的细节。在上例中，当位于洛杉矶的计算机向奥斯丁的计算机发送消息时，这则消息必须经过芝加哥。但是，该消息并不经过公司在芝加哥的所有计算机，准确地说，它仅仅经过那台专用计算机，该计算机是广域网的一个部分。从公司的角度来看，即使全部位于芝加哥的计算机都失效，公司在洛杉矶的计算机仍然能与公司在奥斯丁的计算机进行通信。

8.6 少量广域网与大量局域网

虽然广域网技术已经应用了许多年，但是广域网的成本远高于局域网。我们现在来做一下比较，首先考察局域网的费用情况。一家公司通过在单一建筑物内安装一些电缆来构建连接本公司计算机的局域网。此外，一旦电缆部署到位，这家公司就能比较容易地增添一台计算机到局域网中，即是说，按需扩充局域网。

与低成本的局域网不同，广域网需要更多的设计，而且也需要更多的硬件设施。要部署一个广域网，公司除了必须购置调制解调器、专用硬件设备以及广域网使用的软件外，还必须租用若干长距离传输线路。当然，公司还需要装置一些将公司计算机接入广域网的接口硬件。由于构建广域网的传输线路和硬件设备的价格都十分昂贵，所以只有少数公司才有实力搭建自己的广域网，因此，局域网比广域网流行得多。

8.7 广域网与局域网互不兼容

工程师们已经设计出了各种各样的广域网技术。每种技术在可靠性、速度、覆盖范围和成本等方面都有独立的设计思想和选择。此外，每种技术都选择了符合自身要求的编码技术。因此，大部分广域网技术在电气特性上都互不兼容，人们也不可能通过用一种类型的广域网线路接入另一种类型的广域网中来组建一个更大的网络。

使用广域网的公司常常选择单一的广域网技术来连接公司的所有站点，因此广域网间的不兼容性在公司内部往往并不是什么严重的问题。然而，由于广域网与局域网在电气特性上不兼容，所以这可能引发一个严重的问题。大部分通过广域网连接自身所有站点的公司也使用局域网连接各个站点内的计算机。然而，不兼容意味着公司的广域网不能与公司的任何一个局域网直接相连。

总之：

> 目前存在许多局域网和广域网技术，其中大部分是相互不兼容的。人们无法通过互连两个不同网络来构建一个可用的大型网络。

8.8 单一网络的愿望

使用过多台连接到两个或两个以上不同网络的计算机的人都知道这种方案是多么的不方便。例如，一台计算机接入局域网，另一台接入广域网，那么这两台计算机都无法访问到该团体的所有资源。接入广域网的计算机可以访问远程主机上的资源和信息，却不容易访问本地主机上的信息。相反，接入局域网的计算机可以访问本地资源和信息，但不能访问远程资源。一些需要本地访问的任务对于连入广域网的计算机来说是相当困难或者根本不可能的（例如，用本地打印机打印输出）。同样，需要访问远程资源的任务对于仅接入局域网的机器来说是不可能的（例如，访问远程数据库）。这种方案导致的结果可能是，在一些团体中，需要访问本地和远程资源的用户必须在他的电脑桌上安置两个键盘和两个显示器。在这两台计算机间传送数据就麻烦又耗时。

一个团体中的每台计算机都能访问所有资源，这是最佳情况。当然，流行的策略可能会限制某台计算机或某位用户访问这些资源，而且系统也努力保障资源的安全性，防止非授权用户非法访问。然而，用人工方法把数据从一个网络转移到另一个网络是不必要的。此外，仅仅因为访问不同的资源而迫使用户使用另一台计算机也是很不必要的。

8.9 美国国防部拥有多个网络

20世纪60年代末，美国国防部开始对计算机网络产生兴趣。由于计算机网络的思想是新生

的，所以他们对如何构建网络和如何使用网络几乎一无所知。在军方的资助下，美国国防部高级研究计划署（Defense Advanced Research Projects Agency，DARPA）[1] 开始着手研究采用各种技术实现联网。到 20 世纪 70 年代后期，ARPA 已经拥有了几个可供使用的计算机网络，并开始将这些技术投放到军事领域。ARPA 工程除了包括几个使用卫星和无线电传送信号的通信网络外，还包括一个称作 ARPANET 的广域网。

ARPA 认为军方会面临一个与其有多个网络系统的组织遇到的一样的问题：每个网络都连接了一批计算机，但是不同网络中的计算机之间却没有任何通信链路。从本质上讲，每个网络都形成了一个孤岛，岛中的计算机互连，但岛间却没有任何连接路径。

8.10 连接分散的计算机

ARPA 研究探讨了如何互连一个大型组织的所有机器的问题。ARPA 从基本思想开始，授予工业界和学术界的研究人员研究的许可权，并将研究人员组织起来，共同协作以解决这一问题。研究人员在定期的研讨会上讨论他们的新发现并形成一些新思想。

ARPA 不是让研究人员仅做理论上的研究，而是鼓励他们把自己的思想应用于真正的计算机中。ARPA 选择那些对实验室工作感兴趣的研究人员，坚决要求让他们编写原型软件来验证他们的思想。

8.11 Internet 的出现

ARPA 研究的关键思想是建立一种互连局域网和广域网的新方法，这种方法称作互联网（internetwork，internet）[2]。该术语既表示这个工程也表示根据这个方案建立的系统网络。为了把他们的互联网与其他互联网区分开来，从事 ARPA 工程的研究人员用“internet”表示一般互联网，而用带有大写字母“I”的“Internet”表示他们实验原型的互联网。这个约定保持至今，其关键思想是：

> ARPA 资助研究人员寻求网络间互不兼容问题的解决方案。该工程及研究人员构建的原型系统统称为 Internet。

8.12 ARPANET 主干网

ARPANET 对 Internet 工程来说是特别重要的。由于 ARPANET 是将所有研究人员联系在一起的中心广域网，所以它常常被称作主干（backbone）网络。参与 Internet 工程的每个研究人员都有一台计算机连入到 ARPANET。

只要拥有一个合适的广域网，研究人员们之间就能相互通信，但由于 ARPANET 允许研究员们在一个站点上连接多台计算机，所以 ARPANET 变成了 Internet 工程的关键部分。研究人员利用这一特征将 ARPANET 应用于两个方面。第一，他们像使用传统的广域网一样用 ARPANET 在每个站点上连接一台计算机。第二，他们在每个站点添加一条附加连接，利用这条附加连接来试验他们的新想法。因此，ARPANET 既可以作为标准网络，使研究人员能在各个站点之间传输与工程相关的数据，也可以作为实验性网络，使研究人员能够在其上测试和评估新的网络软件和应用。

---

[1] DARPA 在 20 世纪 90 年代末 00 年代初称作 ARPA（Advanced Research Projects Agency）。

[2] 第 13 章将详细讨论互联网的结构，以及 Internet 是如何连接多个网络的。

8.13　Internet 软件

计算机软件是网络互联技术的重要组成部分[1]。ARPA 主持的 Internet 工程在联网方面做出了许多革新使网络更广泛更有效。研究人员一般独立工作，然后一起设计、测试、改进新的计算机通信方法。他们研究出来软件，使计算机间的通信成为可能。

虽然一个软件由许多相互作用的程序组成，但是研究人员希望这些软件能成为一个统一的系统。他们研究了各类程序之间的相互作用，以此确保一个程序的操作不会跟其他程序产生冲突。最终的结果是形成了一套成熟的、看上去几乎完美的软件。软件的各个部分能很好地协调工作，使大多数用户看不出其内部的复杂性。

8.14　TCP/IP

Internet 软件中有两款软件具有重要的意义。其中，网际协议（Internet Protocol，IP）软件提供基本的通信服务，而传输控制协议（Transmission Control Protocol，TCP）软件提供应用程序所需的附加工具。在不严格讨论中，研究人员把全套 Internet 通信软件用这两个重要成员的首字母来表示，该术语常常写作 TCP/IP[2]。

当需要使用更加正式的名称描述该软件规范时，研究人员称其为 TCP/IP Internet 协议族（The TCP/IP Internet Protocol Suite）。这种正式的名称更加精确，因为它指出整个 Internet 协议族不仅仅只有两种协议。然而，TCP/IP，这个较简单的名称，最终得到了广泛的应用。销售该技术的厂商和获得并安装该技术的用户都使用 TCP/IP 这一术语。

8.15　开放式系统的影响

为了鼓励各家厂商都采用 Internet 技术，ARPA 决定公开其研究结果。当一名研究员发现了一项新技术，或者测量网络性能，或者扩充 TCP/IP 软件，ARPA 要求研究人员将结果记录在文档报告中。编制 TCP/IP 的所有规范和标准以及所有安装和使用 TCP/IP 的经验都需要记录到文档中。ARPA 允许所有人使用这些文档报告。

ARPA 公开网络规范的举动令人震惊，因为大部分设计网络技术的商业公司都会对自己的技术进行保密。实际上，许多公司都对自己的技术申请了专利，以确保没有其他公司使用相同技术。由标准商业行为产生的思想是：

> 流行的商业观点是，销售计算机网络的公司可通过专利来保护他们自己技术，从中可以获得最大利益。

这种思想似乎很有道理，毕竟，如果一家厂商只允许自己品牌的计算机接入自己品牌的网络，其销售额应该会有所增加。该厂商只需要说服一个组织使用它的网络技术，那么该组织就不得不购买它的计算机。

20 世纪 70 年代中期，销售网络系统的主要计算机公司只提供能与它们自己计算机相连的接口硬件。这些技术包括各种局域网和广域网的组合。计算机专业人员在自己的专用系统应用了封闭（closed）式技术，这就表示这些系统对其他系统是封闭的，比如它们排斥其他厂商生产的计算机。

Internet 工程从开始就希望建立一种允许所有厂商生产的计算机都能相互通信的开放（open）式系统。这种开放思想意味着研究人员会公布所有有关 Internet 的发现，并公布编制

[1] 第 15 章到第 18 章将讨论使网络互联成为可能并且高效的各种软件。

[2] TCP/IP 引读法是依次拼出“T-C-P-I-P”。

TCP/IP软件所需的全部规范。

> 如果一家公司拥有某种网络技术，并且使用专利和商业保密措施来防止其他公司制造使用该技术的产品，那么这种网络技术就是封闭式的。相比之下，Internet是一个开放式系统，因为它的所有规范和标准都公之于众，并且任何公司都可以建立与之兼容的技术。

8.16 开放式系统的必要性

计算机公司发现，不管他们多么努力推销自己的封闭式系统，用户仍然会购买几种品牌的计算机。处理器和存储器硬件方面的发展使设计新型计算机成为可能。个人计算机价格的暴跌使普通百姓就能购买得起。像美军这样的组织认识到随着计算机技术的进步，生产厂商将不断生产出新型计算机。此外，并非所有软件都可以在所有计算机上运行。因为一个大型组织需要多种软件系统和计算机来完成多种应用，所以它通常拥有多种品牌和型号的计算机。因此，只有开放式网络系统可以用于不同厂商计算机间的互连。总之，

> 由于一个大型公司常常拥有多家厂商的计算机，所以它需要一个开放式的网络系统。使用封闭式网络系统将限制接入该网络的计算机的种类。

8.17 TCP/IP联机文档

ARPA为Internet工程挑选的大部分研究人员都有使用计算机网络的经历。研究人员来设计和构建ARPANET，并设计其上的诸多应用。研究人员们都清楚他们可以用ARPANET交换技术信息。之后不久，Internet工程启动，他们决定将所有技术文档保存在可通过APRANET访问的计算机文件中。

起初，研究人员计划通过两步来发布报告。当一份报告初稿完成后，作者交由其他研究人员进行评估。一段时间之后，作者综合所有评论意见，并发布该报告的最终版本。要实现这两个步骤，研究人员建立了两组系列文档报告：请求注解（Requests For Comments，RFC）和Internet工程记录（Internet Engineering Note，IEN）。

遗憾的是，这项条理清晰的计划经常会使人误解。研究人员发现，有些最初的报告本身就是完善的，不需要进一步的修订和改定。而其他一些报告被完全重写了一遍，不过为了以后的进一步评论，它们仍然被发布成RFC文档。大部分研究人员发现，不断研究新思想比重复编辑旧文档效果更好。结果，RFC报告成为了Internet工程的官方报告，而IEN被淘汰出局。具有讽刺意味的是，史上最大最成功的计算机网络的技术文档都带有一个工作未完成，笔者仍在等待进一步评论的地方。总之，

> 由于历史的原因，定义TCP/IP及其相关的Internet技术的文档被称作RFC文档。

参与Internet工程的研究人员都能够访问所有的RFC文档，因为这些文档被保存在一台连入ARPANET的计算机中。每个RFC文档都指派了一个整数，并在一张列有数字的索引表上记录了一个索引项。在任何时候，想要了解某个Internet软件细节的研究者都可以通过ARPANET访问包含这些信息的RFC文档。如果该研究人员不记得需要哪些RFC文档，那么他可以通过索引进行检索。

能够在线访问所有工程文档使每个参与Internet工程的研究人员都可以协调他们的行动，并使软件随标准进行更新。更加重要的是，研究人员间的快速通信增加了该工程的进展速度。

> 因为记录有TCP/IP和Internet工程的技术细节的RFC文档可以通过ARPA-

NET 访问，所以 Internet 工程的进度得到很大程度的加速。

随着 Internet 工程的不断推进，技术达到采用和测试原型软件的阶段，于是 Internet 的雏形诞生了。为新生的 Internet 设计的第一批应用之一就是一种用于访问 RFC 文档的应用机制。实际上，几乎所有初期的 Internet 应用都是为工程研究人员提供某种形式的通信。

8.18 军事部门采用 TCP/IP

到 1982 年，Internet 原型已经逐步到位，TCP/IP 技术也已完成测试。几十所学术和工业研究单位一直在应用 TCP/IP，美军也开始在其网络中使用 TCP/IP 技术。

1983 年初，ARPA 使 Internet 覆盖了所有原先接入 ARPANET 的军事站点，这标志着 Internet 开始由实验性网络向实用性网络转折。

8.19 小结

Internet 起初是由 ARPA 资助的研究项目，研究人员研究了多种使用各类网络互联计算机的方法。Internet 这一名称既指该研究项目也指研究人员构建的原型网路系统。

Internet 上运行的称作 TCP/IP 的软件包含许多协同工作以提供通信的复杂计算机程序。该软件运行起来非常好，它隐藏了底层硬件细节，形成了一个看不出接缝的软件系统。

Internet 是个开放式系统，因为任何人可以得到编制 TCP/IP 软件或使用 Internet 所需的标准和规范。设计 Internet 的研究人员以一系列描述 Internet 及其 TCP/IP 软件的报告的形式将 Internet 技术信息公之于众。由于历史的原因，该系列中的每个技术文档都被标记成 RFC。

8.20 个人见解

20 世纪 80 年代末，我参加了一个技术会议，在乘坐电梯的时候有两个与会者也挤了进来，在电梯中，他们进行了一次声音洪亮且生动的对话，其中一人向另一人解释，Internet 作为一项秘密的政府计划已经启动了，目的是升级"热线"（热线就是冷战时期美苏之间开通的专用电话系统）。我本来想打断他们的对话，但我又犹豫了，那个说话最多的人又向电梯里的所有人解释道，当国会听说这件事情后便决定调查该计划的费用，因此军方不得不对这个计划保密。他暗示军方已经和一家大公司达成协议，他还说这家公司为该计划提供资金，作为交换，公司将获得此项技术。我想打断并加以解释，但是这个与会者又以洞悉的口气做出结论，说这条消息显然能解释为什么该公司的股票在去年暴跌，这两个人的谈话引起了众人的讨论和一致的认可。

当步出电梯的时候，我感到既惊奇又迷惑。后来，我意识到有重大的事件已经发生了。Internet 已经迅速变成一种强大的力量。对于新用户来说，它似乎是一个巨大的、陌生的和胁迫的事物。它极大地引起了新用户的注意，并使大家很容易相信包括谣言之类的几乎所有事情，而其他人很容易就能编制解释这些事情的神话故事。这些事情不久就变成了城市里大街小巷的传言。

进修

许多团体都在尝试编写 Internet 的发展史，但是遗憾的是，要想准确的回顾 Internet 发展史并非易事，因为人们的记忆在逐渐淡去，而且每个人的兴趣也与其收集的信息有关。因此，许多网页都包含华而不实或言过其实的关于谁发起了 Internet 或谁跟早期工作有关的言论。

有人试图编写一部准确的 Internet 发展史，我们可以在下面这个网站找到相关信息：

www.netvalley.info

作为一个"打火机"，玩另类，要看爆笑哄骗的互联网历史可以在下面这个网站找到：

http://dogma.burningman.com/~jeremyab/lolih

第 9 章　难以置信的 20 年发展期

9.1　引言

从 1980 年到 2000 年的 20 年间，Internet 从一个小型的、实验性的研究工程发展成为当今世界上最大的计算机网络。到 1981 年，Internet 连接了研究机构和大学院校的近 100 台计算机。20 年后，有超过 6 千万台计算机接入了 Internet。本章将讲述 Internet 的发展历程及其在成长过程中的变化。最后，本章阐述了 Internet 难以置信的发展速度所带来的种种结果。

9.2　传播 Internet 软件

在 1980 年，Internet 仅仅是个研究性工程，但是到了 1985 年，它开始转变成一个可实际运行的网络系统。实验性的 TCP/IP 软件可在几种品牌的计算机上运行。一批大学和科研实验室分别都有一份 TCP/IP 软件，而且每天都在使用它。Internet 已经进入了许多学术和工业研究实验室。

然而，在美国军方能够使用 Internet 进行日常工作之前，这项技术还需要进一步的改善，健壮性应该变得更强，Internet 软件也需要不断测试和完善，整个系统需要进一步的调试。ARPA 仔细考虑了下一步研究计划。

9.3　同时期计算机科学的支持

当 ARPA 从事 Internet 研究工程时，来自一个研究室的另一项技术席卷了整个计算机科学领域，这个技术就是 UNIX 操作系统（UNIX Operating System）。尽管生产厂商将个人计算机内简单的支持软件称作操作系统（operating system），计算机科学家则用这个术语来描述多用户计算机上管理计算机、控制 I/O 设备和提供文件存储的复杂软件系统。由于操作系统的结构十分复杂，所以科学家和工程师们在 20 世纪 60 年代花了几年的工夫去研究它们。各个计算机厂商都编制用于自己品牌计算机的专用操作系统。

贝尔实验室的一组计算机科学家在 20 世纪 70 年代初编制了一种新型操作系统，他们把它称作 UNIX 分时系统（UNIX Timesharing System）。因为贝尔实验室使用了多种不同的计算机，所有研究员们编制的这套系统是通用的，即是说，他们认真地设计这套软件，以便使其容易移植到新的计算机上。

贝尔实验室允许各所大学把 UNIX 系统用在教学和科研上。由于贝尔实验室想要评测 UNIX 系统的可移植性，所以它向各所大学提供了源代码，并鼓励它们尝试在新机器上运行该系统。因此，UNIX 系统成了学生们可以学习研究的第一批操作系统之一。

来自加州大学伯克利分校的一批有才华的研究生开始对 UNIX 系统产生兴趣。他们编写了基于该系统的应用程序，同时也修改了系统本身。他们给系统添加了一些新特性，并实验了使用局域网的程序。为了使这些成果可为其他大学所用，伯克利分校的研究者们建立了一套软件发布工具，以便邮寄包含他们软件的磁带。于是，UNIX 系统的伯克利版本诞生了，它通常被称作

BSD UNIX[1]，很快这套系统在其他高校流行起来。

9.4 Internet 适合 UNIX

ARPA 认识到伯克利软件已经分发到多所大学，于是决定利用它来散播 Internet 软件。ARPA 与伯克利方面达成了一份研究协议，在协议的条款中，ARPA 将作为 Internet 工程一部分而研发的 TCP/IP 软件提供给伯克利方面的研究人员。伯克利方面把 TCP/IP 软件集成到他们自己的 UNIX 系统中，并且修改了要用到 TCP/IP 的应用程序。

当伯克利方面发布了它若干一个主要软件分发系统后，大部分计算机科学系都可以免费获得 TCP/IP 软件。虽然只有少数计算机科学系有计算机连入 Internet，但是更多的系都已有或即将安装一个局域网。学校知道他们的学生都需要学习网络知识，他们也知道使用网络将使计算更加容易，因为网络允许用户共享诸如打印机之类的资源。

对大多数部门来说，TCP/IP 是它们用过的第一款实用的网络软件。TCP/IP 提供给部门级网络一种费用低、效率高的组网方法，也是一种可以用于课堂探讨的技术。因此，不久之后，大部分计算机科学系都在自己的局域网中运行 TCP/IP 软件。要点总结如下：

> 高校的计算机科学系在免费获得加州大学伯克利分校的 UNIX 系统软件的同时，也免费得到了 TCP/IP 软件。虽然只有少数计算机系有计算机接入 Internet，但是大部分计算机系都在自己的用于教学、科研和高性能计算的局域网系统中使用 TCP/IP。

9.5 美国军方的认可

到 20 世纪 80 年代初，Internet 已经能够稳定地运行了。它连接了许多学术和科研站点。更为重要的是，Internet 已经证明互联网的基本原理是健全的。美国军方在确认了 Internet 的可行性后，也把本部门的计算机接入到 Internet，并开始使用 TCP/IP 软件。

1982 年，美国军方将 Internet 作为主要计算机通信系统，接着，又确定了最后期限。1983 年初，ARPANET 和相关军事网络中正运行旧的通信软件，所有连接技术都转而采用 TCP/IP。任何不支持 TCP/IP 的计算机都无法进行通信。要点总结如下：

> 虽然美国军方资助了 Internet 研究项目，并且最终选择了使用 Internet，但是互联网技术是在民用部门开发和测试的。

9.6 Internet 的规模在一年内扩大了一倍

在美国军方将他们的计算机转换到 TCP/IP 之前，Internet 互联了近两百台计算机。一年之后，Internet 规模扩大了一倍。现在看来，这个增长是微不足道的，因为这仅仅只有数百台，还不到上千台计算机。然而在那个时刻，这种增长还是具有相当重大的意义。

人们容易想到，Internet 规模的扩大暴露了计算机软件的局限性。例如，TCP/IP 中的一部分包含有其他计算机和用于访问它们的地址的列表。随着接入 Internet 的计算机的增多，这张表变得过大，以至于不得不修改软件以容纳较长的列表。与现代计算机不同，20 世纪 80 年代初期的计算机的内存都比较小，研究人员不可能任意地增加内存容量。由于该列表所占据的内存不能为网络设备或其他应用程序所用，所以，列表占据的内存越多，计算机的运行速度就越慢。

起初，研究人员对软件功能进行了少量的扩充，他们将软件的容量增大 10%或 20%。不久

---

[1] 首字母缩写 BSD 全称是 Berkeley Software Distribution。

之后，他们发现这样做根本就不够，还需要进一步的扩容。随着 Internet 规模继续扩大，软件扩容的步调也必须跟上。

除了暴露出软件的局限性之外，Internet 的增长还显露了人力和办公方面的不足。比如，每当有新的计算机添加到 Internet 中时，就会涉及几个人。一个人在有权连接之前必须先说明连接的理由和这与 Internet 工程的关系，另一个人必须给这台计算机分配一个名字，并将其录入到数据库中。当然，最后还必须有人将这台计算机物理连接到 Internet 中。

在 Internet 飞速发展期间，由于研究者们忙于更新软件，他们几乎没有时间去从事诸如注册之类的手工步骤。于是这些事情就开始交由专门的人员来完成。本节内容总结如下：

> 随着新的计算机的加入，Internet 在一年时间里规模扩大了一倍。这迫使研究人员改进了软件，并调整了管理步骤。

9.7 计算机科学部门

20 世纪 70 年代末，许多计算机科学家都意识到了网络的重要性，一个研究小组向美国国家科学基金会（National Science Foundation，NSF）[1]提交了一项联网计划，其目标是设计一个连接所有计算机科学研究人员的网络。

在对该计划进行评估，并交由该小组修订之后，NSF 对这项建造计算机科学网络的项目进行了投资。该项目也得到了 ARPA 的资助，后来它成了众所周知的 CSNET。

为了覆盖全国的计算机科学家，CSNET 不得不解决给各类研究机构提供网络服务的问题。由于大型研究机构能够负担较高的费用，所以 CSNET 并鼓励它们运行 TCP/IP 软件，并鼓励它们接入 Internet。而对于那些较小的研究机构来说，它们不能担负直接连接服务的费用，因此 CSNET 设计了以较低费用提供有限网络服务的方法。

在美国军方选择 Internet 作为其主要的计算机通信系统之前，许多工业和学术界的顶级计算机科学组织就已经在使用 Internet 了。在接下来的几年中，CSNET 致力于向计算机科学部门提供 Internet 连接服务。结果，到 20 世纪 80 年代中期，多数计算机科学家都能访问 Internet 了。

9.8 研究生自愿付出他们的时间

将主要计算机科学研究组织接入 Internet 产生了一些影响很有意思。虽然部分计算机科学家供职于企业的研究实验室，但他们大多数都是大学教授，他们在讲课的同时，还向学生提供指导。教授们向学生讲授 Internet 工程及其运用的技术和软件、Internet 成功的原因，及其目前尚存的问题。教授们的热情极具感召力。

于是，学生们逐渐开始对学习更多关于 TCP/IP 和 Internet 的知识有兴趣。正在寻找研究课题的研究生们开始探究 TCP/IP 软件的技术细节。他们探索扩充 Internet 技术的各种方式方法，并设计评估这些技术性能的实验方法。研究生还研究各种新应用，并探索扩充网络功能的方法。最终的结果是双赢的，学生获取了计算机网络方面有价值的知识和经验，而他们的创造力有助于改进 Internet 技术。

9.9 IAB 的演变

从事 Internet 研究的科学家们定期召开会议，讨论新思想、评论新技术、分享最新发现，并交换技术信息。ARPA 认为，随着 Internet 的飞速发展，科学家团体应该有更为正式的组织形

---

[1] NSF，美国的一个联邦机构，负责对科学和工程领域的研究和教育提供资金。

式，也应该肩负更大的责任以协调 TCP/IP 的研究和 Internet 发展，于是，ARPA 把这一组织更名为 Internet 活动委员会（Internet Activities Board）。按军事上的惯例，该委员会的名称选取各单词的首字母，简称 IAB。

ARPA 任命了 IAB 的主席，虽然迅速发展的 Internet 不可能由某一个人设计出它的整个体系结构，但是 ARPA 还是授予了 IAB 主席"Internet 建筑师"（Internet Architect）称号。IAB 的另一名成员被任命为"RFC 编辑"（RFC Editor），他负责所有 RFC 文档在发布之前的评估和编辑。IAB 的其他科学家都各自被分配了一项特定的研究课题。

为了研究各自的课题，IAB 的每一名成员都从各自的研究领域召集了一批研究志愿者，并组成一个任务组（task force）。IAB 成员担任各自小组的主席，并与小组成员通力协作，代表小组参加 IAB 会议。每个任务组都会经常召开会议，讨论思想，解决问题，产生新方法，以及报告实验情况。如果一个任务组就一项新方法达成一致，那么成员们就会编制原型软件来演示他们思想的实际运作情况，然后他们会制定并提交一个标准，将这作为 RFC 文档。

IAB 引导 Internet 的发展很多年，1989 年，IAB 实现了重组，新组织吸收了商业机构的代表。1992 年，IAB 成为 Internet 协会（Internet Society）的一个组成部分，此时，它的职责和与其他组织的关系都重新进行了修订。IAB 丢开了一些技术上的职责，把更多的控制权下放到子团体，而自身仅仅作为策略和标准的最终仲裁者。在 IAB 第二次重组的时候，其缩写不变，但其全称变成了 Internet 架构委员会（Internet Architecture Board）。

9.10 IETF

在 IAB 建立的所有任务组中，有一个组显得特别突出，它就是 Internet 工程任务组（Internet Engineering Task Force，IETF）。IETF 在 IAB 重组后保留了下来，并且一直保持活跃。事实上，IETF 已经发展得相当庞大，以至于它的下级组织分布在十几个研究领域，并且每个研究领域都有一个负责协调各组工作的管理人员。IETF 每年大约召开 3 次公开会议，会议地点在欧洲、美国和太平洋地区轮流选择。每次会议都有数百人参加，参会人员大都来自各大商业公司。他们都是自愿参加的，目的是了解最新发展动态，推动软件的修订和改进。

IETF 成立的初衷是解决 Internet 的短期发展问题。现在，它的职责是确定技术发展方向，包括新的通信软件标准或旧软件的修订。如今，大部分 RFC 文档都出自 IETF 中被称作工作组（working group）的委员会。总之：

> 负责引导 Internet 研究和发展的团体是 Internet 体系结构委员会（IAB），而
> IAB 中主要负责技术事务的下级组织是 Internet 工程任务组（IETF）。

9.11 Internet 规模一年内再次翻番

在美国军方决定使用 Internet 之后的几年中，Internet 继续发展。1984 年，Internet 规模再次翻番，但是，仅仅规模并不能反映这样一个事实：ARPA 之外的其他政府部门也开始使用 Internet，并支持该研究计划。例如，美国国防部（Department of Defense，DOD）和美国航空航天管理局（National Aeronautics and Space Administration，NASA）都在各自的部分网络中采用了 TCP/IP，并且不久之后将更多的网络接入 Internet。

9.12 Internet 促进科学发展

到 20 世纪 80 年代中期，NSF 意识到科学的发展很快就会需要计算机通信。在计算机网络出现之前，科学家们通过在科学期刊上发表文章来交流思想。科学家从提交手稿到最终发表需要

几个月，甚至几年的时间。

计算机通信改变了科学家们从事研究的方式。接入 Internet 的科学家们能即时地交换文档或实验数据。实际上，科学家在 Internet 上公布进展中的实验数据，这使得其他科学家无需跑到实验现场就能分析这些实验结果。更为重要的是，科学家可以在 Internet 上进行非正式的不公开的讨论。

9.13 NSF 的领导角色

当认识到 Internet 对科学的重要性后，NSF 决定资助 Internet 的发展和 TCP/IP 技术。1985 年，NSF 宣布他们计划将 100 所高校的研究人员连入 Internet。此外，NSF 还向美国国会提交了一份计划，并获得了一笔资金用于联网。NSF 咨询了网络领域的专家学者，然后制定了一份计划，并开始实施一项使 Internet 发生重大改变的方案。

科学家们经常使用复杂、高速的计算机来分析实验数据，这类计算机称作超级计算机（supercomputer）。因为超级计算机十分昂贵，所以之前 NSF 在美国建设了 5 个超级计算机中心。从事 NSF 项目的科学家都能使用最近的超级计算机中心来分析数据。

NSF 的第一步计划是建立一个互联 5 个超级计算机中心计算机的广域网。这个网络采用 TCP/IP 技术，并提供接入 Internet 的连接。该网络称作 NSFNET，它的规模比 ARPANET 小得多，其速度也远不及 ARPANET。虽然科学家们都认为 NSFNET 有用，但它却没有什么令人兴奋的地方。

9.14 目标：涵盖所有科学和工程研究人员

NSF 知道他们建立的这个小型网络不可能取代 ARPANET，但这仅仅是开始。当科学家开始使用这个初级网络访问超级计算机时，NSF 制定出一项新的重大方案。该方案确立了一个宏伟目标：

> NSF 认为，为了保持美国的竞争力，必须让网络访问覆盖到每门科学和所有工程研究人员。

尽管 Internet 提供的诸多功能给 NSF 留下了深刻的印象，但是同时他们也发现，ARPANET 没有足够的能力完成这个目标。显然，Internet 需要一个新的广域网。

9.15 NSF 的做法

NSF 决定拿出资金建立一个新的 Internet，其容量比现存的 Internet 大得多。在考察了可用技术及审订了预算后，NSF 认为自己无法承担整个工程的经费。作为折中，NSF 决定以联邦授权的形式提供部分资金支持。其他组织及各大公司向 NSF 提交书面建议，请求获得工程所需经费。

NSF 决定将许可权分作两类。第一，NSF 资助那些想要建设和运行连接部分 Internet 的新级高速广域网的团体。新广域网不得不取代 ARPANET 的一部分以及初期的 NSFNET。第二，NSF 资助那些想要互联小范围内计算机并将它们接入新广域网的团体。例如，NSF 认为，每个州可以各自组成一个团体。起初，NSF 把这些团体称作 NSF 地区网络（NSF Regional Network），后来，当有些团体覆盖了更大的地理范围时，NSF 开始把它们称作 NSF 中级网络（NSF Mid-Level Network）。

由于大多数高校或公司已经通过局域网互联了本单位的计算机，所以 NSF 决定用其经费帮助他们进行远程连接，而高校和公司自己承担内部网络费用。

9.16 NSFNET 主干网

NSF 采用竞标的方式来颁发新型 Internet 广域网的许可权，这个新网络称作 NSFNET 主干网（NSFNET backbone）。1987 年，NSF 征求各方建议，然后请一组科学家帮助评估这些建议。在综合考虑多种方案后，NSF 采纳了来自 3 个组织的联合提案，这 3 个组织分别是：计算机制造商——IBM；长途电话公司——MCI；建立和运行连接密歇根州各所学校的网络的机构——MERIT。

这 3 家组织共同建立一个新的广域网，这个广域网在 1988 年成为了 Internet 的主干网。MCI 提供远距离传输线路，IBM 提供广域网中用到的专用计算机和专用软件，而 MERIT 操作整个网络。很多人仍然沿用其前身 NSFNET 来命名这个新的主干网。

9.17 ANS 主干网

最终，当这个新广域网的流量达到容量上限时，NSF 决定对这个网络进行迅速地重组，要将每条传输链路的容量提高到原来的 3 倍。由于 Internet 增长过快，到 1991 年末，NSFNET 主干网很快又达到容量上限。NSF 意识到联邦政府不可能无限制地负担 Internet 的费用，他们希望将部分责任转嫁到企业。为了解决这一问题，IBM、MERIT 和 MCI 联合成立一家非盈利公司——高级网络与服务公司（Advanced Networks and Services，ANS）。

1992 年间，ANS 建立了一个作为 Internet 主干网的新型广域网——ANSNET。这个广域网采用的传输链路容量是它所取代的 NSFNET 的 30 倍。图 9-1 示例了主要的 ANSNET 连接。

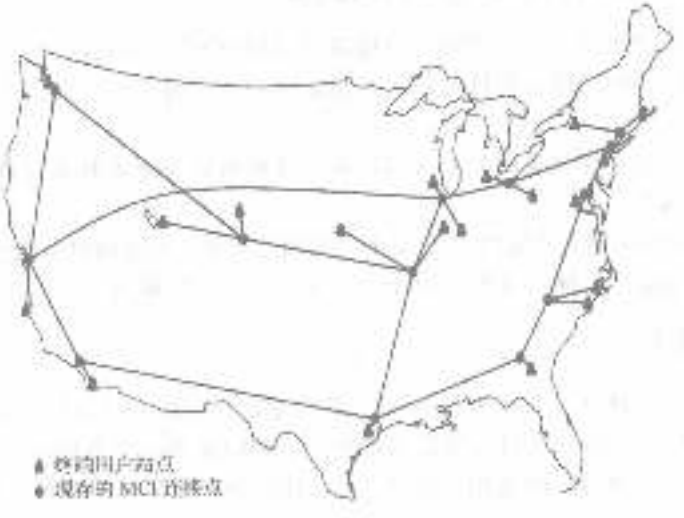


图 9-1 构成 Internet 主干网的广域网（1995 年），此网的建设资金来自 NSF、IBM、MCI 和 MERIT

ANSNET 与之前的 NSFNET 有根本性的区别，ANS 作为非联邦政府部门，它拥有传输链路和联网计算机的所有权。将所有权移交给企业是 Internet 商业化和私有化进程的第一步。

1995 年，MCI 公司开始着手研发更大容量的广域网，这就是后来众所周知的高速主干网络系统（very high-speed Backbone Network System，vBNS）。后来，NSF 提供资金研发名为 Internet2 的主干网。

9.18 Internet 呈指数级增长

NSF 将科学家和工程师组织起来后，Internet 以难以置信的速度增长。1983 年，Internet 共连接了 562 台计算机，然而，10 年后，它连接的计算机超过 120 万台，并且仍然在飞速增长。通过下列数据，我们能够更好的理解这种惊人的速度：

> 到 1999 年，Internet 的增长速度达到平均每秒就有一台计算机进入 Internet；
>
> 到了 2006 年，平均每秒有超过 10 台计算机接入 Internet。

尽管 Internet 并不是每年都以完全相同的速度增长，并且大多数计算机都是在最近几年才接入 Internet 的，但是再次翻番的趋势是必然的。Internet 的增长速度平均保持在每月增长大约 10%，每 10 个月网络规模翻一番。数学家把这种增长称为指数级增长。图 9-2 的表格中列出了自 1983 年到 2006 年的增长情况。

指数级增长具有一些有趣的特性。比如，虽然 Internet 已经运行了这么多年，但指数级增长意味着有近一半的用户是上一年才刚刚连入 Internet 的。有意思的是，每一年的情况都符合这条规律。实际上，我们这样总结这种惊人的增长：

> 从 1983 年到 2005 年的任何时候，Internet 中大约有一半的增长发生在前 12 个月到 14 个月。

理解指数级增长的另一种方式就是看单个人的经历。起初，一个刚开始访问 Internet 的人觉得自己就是一个新手，好像其他人访问 Internet 的时间都很长了。但是一年后，他会发现有超过一半的 Internet 用户刚开始访问 Internet。

将时间尺度扩大一些对理解指数级增长有更好的帮助。如果指数级增长持续下去，那么 3 年后，就有超过 87% 的 Internet 用户是最近才加入的。6 年后，这个数字将达到 98%。2000 年的时候，我们已经认为 Internet 足够庞大了，其实那时它还不到当前规模的 15%。

年份	上计算机的大概数量
1983	562
1984	1,024
1985	1,961
1986	2,308
1987	5,089
1988	28,174
1989	80,000
1990	290,000
1991	500,000
1992	727,000
1993	1,200,000
1994	2,217,000
1995	4,852,000
1996	9,472,000
1997	16,146,000
1998	29,670,000
1999	43,230,000
2000	73,000,000
2001	109,574,429
2002	147,344,723
2003	171,638,297
2004	233,101,481
2005	317,646,084

图 9-2 从 1983 年到 2005 年，每年接入 Internet的计算机数量的估计值。自 2000 年开始，NAT 技术的采用使获取准确的计算机数量变得比较困难。到 2005 年，Internet 上的大多数计算机在数据上都体现不出来[1]

9.19 商业评估

这种增长既有害也有利。一方面，它使负责运行 Internet 的团体感到头疼，也使设计新技术以适应增长的工程师们备感压力。另一方面，Internet 的增长给那些销售 TCP/IP 和 Internet 技术的厂商提供了难以想像的机遇。

---

[1] 第 19 章讨论了 NAT 技术如何允许多台计算机共用一个地址。

在 20 世纪 80 年代中期，TCP/IP 软件主要用于各高校或计算机科研实验室。如今，商业部门和个人用户都开始使用 TCP/IP，一个完全崭新的产业局面出现了，企业销售产品和服务，使家庭和商业部门都能接入 Internet。更为重要的是，有公司开始销售通过 Internet 传送信息的服务。

9.20　Internet 增长的尽头

Internet 不可能无限地增长下去，虽然各种技术部在适应这种惊人的增长，但指数级的增长终究会走到头。比如，按照当前的增速，Internet 的增长将会很快超过全世界计算机的生产速度。此外，当前的 Internet 技术可以连接最多 4 294 967 295 台计算机[⊖]，这是现有计算机总量的好几倍。但是这个数量仅仅是当前世界人口的 2/3。而增速到何时增长会降到指数级增长以下？何时又会停下来呢？

过去很多时候曾多次有人预言 Internet 很快就会瓦解，因为，他们发现该技术的部分环节正接近极限。例如，1990 年那年，有人预测 Internet 到 1993 年 3 月将不复存在。1995 年，又一些人预言它会在 2004 年崩溃。但事实证明，这些毁灭性的预言都破灭了，Internet 仍在成长。每当网络流量接近主干网容量时，一种新的主干网络技术就会出现，使容量大大增加。当流量接近组成 Internet 的专用交换计算机的交换能力时，更快的计算机就会应运而生，产生更强大的处理能力。当研究人员一致认为 Internet 的增长必须加以限制时，没人能够预测到它的增长何时会慢下来。

进修

Internet 系统财团通过域名系统（将在第 18 章讲解）统计过注册域名的计算机的数量。关于这方面的信息在下面的网址可以找到：

http://www.isc.org/index.pl?/ops/ds

除了计量 Internet 中计算机的数量之外，有几个组织正尝试估算访问 Internet 的人数。例如，按地域统计的 Internet 用户数在下面的网址中可以找到：

http://www.internetworldstats.com/stats.htm

⊖ 产生这种限制的原因在于 TCP/IP 用来标识计算机的数字地址必须小于 4 294 967 295。新技术允许多台计算机共用一个地址，而 IP 协议的最新版本则打破了这一限制。

第 10 章　全球化的 Internet

10.1　引言

因为 Internet 跨越了许多国家，所以人们常把 Internet 称作全球 Internet（The Global Internet）。本章将介绍 Internet 的覆盖范围，并举例说明它是如何扩展到世界各地的。

10.2　早期的 ARPA 网络

尽管由 ARPA 资助的大多数早期的网络研究都集中在美国，但是其中仍然有部分网络的研究与其他国家有关联。例如，ARPA 通过卫星与挪威和英格兰的站点进行实验性通信。

在 ARPA 重点支持 Internet 项目时，它用美国之外的已有的连接来对 Internet 技术进行更大规模的测试。因此，这些站点就成了第一批访问 Internet 的国外站点。

10.3　计算机间传送的电子邮件

在 ARPA 的研究人员从事 Internet 研究的同时，美国其他的研究者们也在做联网的实验。这些研究人员中有许多都使用电子邮件与同一分时主机中的其他用户通信。他们认识到，如果一台计算机可以进行基本的通信，那么其电子邮件系统也能向其他计算机上的用户发送消息。这些研究者们编写了使用拨号电话系统连接计算机的软件，之后不久，他们就在互联的多台计算机上架设了电子邮件系统。

将电子邮件连接扩展到美国之外的其他国家的计算机上是一件比较容易的事。因为语音电话网络采用统一的标准，所以各国的电话系统都可以互换作。因此，研究者们无需增加额外的硬件就可以实现通信。与美国一样，其他国家的计算机也只需要一个调制解调器和一份软件就可以实现通信。

伴随着贝尔实验室开发的 UNIX 系统的出现，一种可在拨号电话系统上提供电子邮件服务的早期计算机网络技术也应运而生。该技术就是称作 UNIX 到 UNIX 的拷贝程序（UNIX to UNIX Copy Program，UUCP）的应用程序。它处理计算机和调制解调器之间的交互细节，拨打目的计算机的电话号码或接收呼入的拨叫，并在建立的电话连接上传输电子邮件。当计算机所有者同意交换电子邮件时，UUCP "网络" 也随之形成了。

10.4　BITNET 与 FIDONET

并非人人都安装 UNIX 系统。使用 IBM 大型机的研究人员们发明了一种允许那些没有安装 UNIX 的计算机交换电子邮件的网络，人们把这种网络称作 BITNET。BITNET 是从底层实现的，它与 UUCP 具有根本的差别。还有一种称为 FIDONET 的网络也允许计算机发送电子邮件。许多国家采纳了像 UUCP、BITNET 和 FIDONET 这样的技术，这些技术一直到 20 世纪 90 年代末还在使用。例如，直到 1997 年，仍然有一些国家没有接入 Internet，它们都使用 FIDONET 收发电子邮件，这些国家是：

乍得、阿比亚、马拉维、塞拉利昂

图 10.1 列出了到 1997 年为止所有没有接入 Internet 的国家，但是这些国家能相互收发电子邮件并能和 Internet 站点进行电子邮件交互。

安哥拉	格林纳达	萨摩亚群岛
安圭拉岛	瓜德罗普岛	塞舌尔
巴哈马	几内亚	塞拉里昂
孟加拉国	圭亚那	所罗门群岛
波斯尼亚和黑塞哥维纳	海地	圣文森特和格林纳丁斯
博茨瓦纳	基里巴斯	苏丹
布基纳法索	老挝	斯威士兰
柬埔寨	莱索托	塔吉克斯坦
喀麦隆	马拉维	坦桑尼亚
乍得	马里	多哥
库克群岛	马绍尔群岛	汤加
科特迪瓦	毛里	土库曼斯坦
古巴	荷属安的列斯群岛	图瓦卢
塞浦特岛	新克里多尼亚	瓦希阿图
多米尼克	尼日尔	越南
法属圭亚那	尼日尼亚	南斯拉夫
法属波利尼西亚	纽埃岛	
冈比亚	巴布亚新几内亚	

图 10.1　1997 年能够通过计算机网络收发电子邮件的国家，但它们并没有直接连入 Internet

10.5　欧洲的网络

除了计算机技术的出现，欧洲也开始建立自己的计算机网络。欧洲的大部分国家都有一个名为邮电总局（Post，Telegraph，and Telephone，PTT）的机构。该机构隶属于政府部门，负责控制包括计算机网络在内的各种形式的通信。

PTT 曾与其他电话公司一起合作过，为了确保世界上的所有电话系统相互兼容，它们组建了一个制定标准的组织，该组织的官方名称是国际电信联盟（International Telecommunication Union，ITU）。ITU 发布各种包含电话系统技术标准的文件。例如，这些文件指定了诸如电话线上的电压值的细节，以及电话号码分配的国际标准。

当 PTT 开始对计算机网络发生兴趣时，它们要求 ITU 制定一项确保兼容性的网络标准。ITU 成立了一个专门的委员会来制定这项计算机网络标准。结果产生了一种名为 X.25 的网络技术，该技术在整个欧洲得到广泛使用。

由于欧洲各国的 PTT 都使用了 ITU 制定的标准，所以欧洲各团体要试验其他技术（如 Internet 上采用的 TCP/IP）是很困难的。总结如下：

> 欧洲的多数国家都在计算机网络中采用 X.25 技术，这是因为各国负责控制网络的 PTT 都遵从 ITU 推荐的计算机网络标准，这点就像当年 PTT 遵从 ITU 推荐的电话网络标准一样。

尽管联网上存在限制，但欧洲各高校和研究实验室的研究者们都在设法开发一些实验网络。比如，在英国，一个名为联合学术网络（Joint Academic NETwork，JANET）的网络在 20 世纪 70 年代便开始运行。在 IBM 的资助下，几个欧洲国家的高校建立了一个网络，称作欧洲学术和研究网络（European Academic And Research Network，EARN）。

10.6 EBONE：欧洲的 Internet

1991 年，一些欧洲国家已经在使用基于 TCP/IP 的实验性网络了，此外，部分国家也已接入了 Internet。大多数实验网络都连接了高校和研究实验室的计算机。遍及欧洲的各个团体相互协作，目的是形成一个高速的欧洲主干网络，连接所有团体并将 Internet 延伸到每一个人。图 10.2 给出了最终形成的欧洲主干网络示意图，人们把它称作 EBONE。

到了 20 世纪 90 年代中期，欧洲主干网成员数已达 25 个，每个成员每年都要支付一次年费，相应地，它们能获得与其他站点以及 Internet 美国部分的可靠连接。欧洲主干网中心组织利用这些钱作为维护主干网的费用。例如，该组织不得不租用专用传输线路（包括连接到美国的线路），以及向管理操作主干网的人员支付薪水。总之：

EBONE 是覆盖整个欧洲并与全球 Internet 站点相连的广域网。

10.7 主干网和 Internet 的层次结构

与美国一样，Internet 的欧洲部分的组织结构也分三层。EBONE 处于顶层，连接欧洲各个地区。此外，一个地区有一个或多个网络，这构成了层次结构中的第二层。多个站点接入地区网络，再由地区网络接入主干网。在第三级，各个站点通常都包括多个局域网，这些局域网连接了站点里的多台计算机。

Internet 仍然在使用层次结构。当位于同一站点里的两台计算机需要通信时，它们直接使用站点所在的本地网络。位于同一地区不同站点里的计算机通信依赖的是连着这些站点的地区网络。最后，当一个地区的计算机与另一个地区的计算机通信时，数据通过主干网从一个地区传到另一个地区。

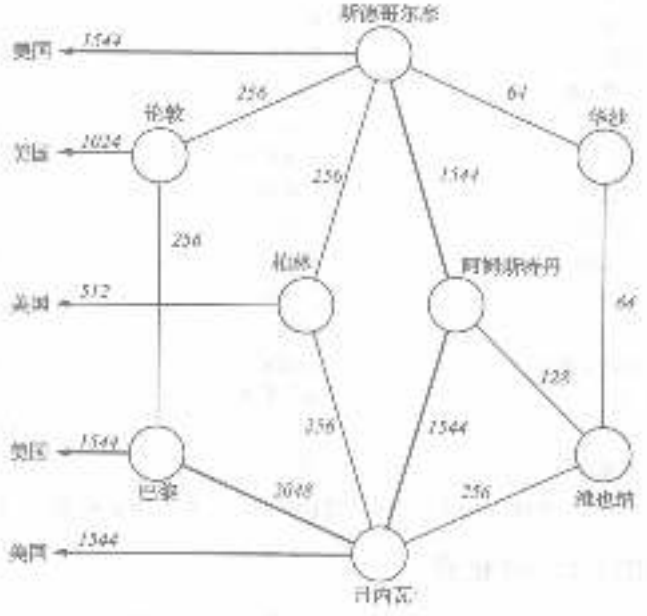


图 10.2 欧洲主干网 EBONE 及其连接到 Internet 美国部分的传输线路示意图（1995 年）。每条传输线路都标有容量（单位：kbps）。到 2000 年，商业服务业界开发出了另一个主干网，EBONE 就逐渐变得过时了

10.8　各大洲的 Internet

到 1997 年，Internet 已经覆盖到了世界上包括南极洲在内的七大洲，图 10-3 列出了接入 Internet 的 172 个国家和地区。

阿尔巴尼亚	多米尼加共和国	吉尔吉斯斯坦共和国	俄罗斯联邦共和国
阿尔及利亚	厄瓜多尔	拉脱维亚	圣马力诺
安道尔共和国	埃及	黎巴嫩	沙特阿拉伯
南极洲	萨尔瓦多	列支敦士登	塞内加尔
安提瓜和巴布达	爱沙尼亚	立陶宛	新加坡
阿根廷	非罗岛	卢森堡	斯洛文尼亚
亚美尼亚	斐济	中华人民共和国澳门特别行政区	南非
阿鲁巴岛	芬兰	马其顿	西班牙
澳大利亚	法国	马达加斯加	斯里兰卡
奥地利	法属圭亚那	马来西亚	圣卢西亚
阿塞拜疆	德国	马耳他	苏里南
巴林	加纳	毛里求斯	斯瓦尔巴岛和扬马延岛
巴巴多斯岛	直布罗陀	墨西哥	瑞典
白俄罗斯	希腊	摩尔多瓦	瑞士
比利时	格林兰	摩纳哥	中国台湾
贝宁	关岛	蒙古	泰国
玻宁	危地马拉	摩洛哥	特立尼达和多巴哥
百慕大	洪都拉斯	莫桑比克	突尼斯
玻利维亚	中华人民共和国香港特别行政区	纳米比亚	土耳其
巴西	匈牙利	尼泊尔	乌干达
文莱达鲁萨兰国	冰岛	荷兰	乌克兰
保加利亚	印度	新西兰	阿拉伯联合酋长国
加拿大	印度尼西亚	尼加拉瓜	英国
开曼群岛	伊朗	挪威	美国
中非共和国	爱尔兰	巴基斯坦	乌拉圭
智利	以色列	巴拿马	马兹别克斯坦
中国	意大利	巴拉圭	梵蒂冈
哥伦比亚	牙买加	秘鲁	委内瑞拉
哥斯达黎加	日本	菲律宾	维尔京群岛
克罗地亚	约旦	波兰	赞比亚
塞浦路斯	哈萨克斯坦	葡萄牙	津巴布韦
捷克共和国	肯尼亚	波多黎各	
丹麦	韩国	留尼旺	
吉布提	科威特	罗马尼亚	

图 10-3　1997 年接入全球 Internet 的国家或地区，之后所有国家或地区都接入了 Internet

10.9　1998 年后的 Internet 世界

有趣的是，1998 年是 Internet 发展的又一里程碑。1998 年间，Internet 连接了世界上的所有国家。当然，有一些国家仅有少量计算机接入 Internet，而很多国家的连接速度都比较慢，尽管如此，Internet 仍然广受欢迎。

到 1998 年底，Internet 覆盖了世界上每一个国家。

10.10　个人见解

随着 Internet 的发展，我收到了来自世界各地的消息。20 世纪 80 年代初，我对收到来自英格兰和加拿大的消息感到兴奋异常，因为一想到网络连接已经到达了其他国家就令人十分激动。当我第一次收到一封来自苏联的电子邮件时，我感到十分惊讶，毕竟在冷战时期这样的通信是被禁止的。当我第一次收到来自远方的信件时，比如澳大利亚、日本、印度、中国，同样觉得十分高兴。

尽管与其他国家通信的兴奋感逐年消退，我仍然会为 Internet 巨大的渗透力感到震惊。加利福尼亚的公交服务系统通过提供无线 Internet 连接服务来吸引人们使用其设备。显然我个人希望商务酒店提供 Internet 接入服务，但我惊讶地发现旅游圣地的一些小旅馆也提供了高速 Internet 服务。其实，我本不应该这样惊奇，毕竟，Internet 是全球性的网络。

第 11 章　全球的信息基础设施

11.1　引言

上一章介绍了构成 Internet 的远程连接中的一部分，并说明了 Internet 是如何扩展到全世界的。Internet 的重要性不仅仅在于其庞大的规模，还在于它所提供的全球性的通信结构，正在改变人们交流的方式。本章首先回顾以往的基础设施发展是如何改变社会的，然后说明 Internet 潜在的影响力。

11.2　现存的基础设施

最初，术语基础设施（infrastructure）指的是军方建立的永久性基地，通常是用于防卫的防御工事。现在，它是指社会依赖的公共服务系统。例如，向家庭和商业供电的电力系统就是工业国家基础设施的重要组成部分。

每种新基础设施都改变了社会。在货币出现之前，所有的商业交易都是物物交换。有物品或服务要出售的人们不得不与其他物品或服务交换。每次交易都需要双方协商具体的交换事项。货币引入了价格这一概念，使人们可以把物品或服务的价值转换成一个通用的形式。

陆上运输基础设施改变了商业运作的方式。在建立陆上运输设施之前，各公司的厂址只能选择在靠近原料产地、工人和客户的地方。铁路的出现改变了社会，这是因为它们使长途运输原材料或产品经济得多。铁路运输基础设施的建立意味着公司可以在远离原料产地或客户居住地的地方修建工厂、生产商品。

美国交通运输设施的第二次大转变发生于美国政府建立横跨全国的州际公路系统这一时期。最初，设计这套公路系统是用于输送军事装备。州际公路系统包括连接各州的高速公路，这使人们迅速抵达本国的任何地方成为可能。虽然铁路先于公路出现，但州际公路系统的建立增加了最基本的基础设施，因为它使每个人都可以灵活地选择目的地以及直接到达目的地的路径。结果，公路长途运输比铁路更加经济，于是卡车工业迅速得到发展。总结如下：

> 新的基础设施促进了新型工业的发展，州际公路系统就是一个例子。

11.3　通信基础设施

通信设施的发展取决于多方面的进步。在古文明中，信使靠双脚送信，后来，信使骑马送信。现代文明下的邮政服务系统取代信使完成送信任务，伴随邮政服务系统建立的基础设施具有两方面的先进之处。其一，因为邮政服务允许邮件发送给个人，这意味着任何享受邮政服务的个人都能收到信件。这种思想演变成了众人皆知的普遍投递（universal delivery）。其二，现代的大多数邮政系统都允许任何人邮寄信件，通信不再局限于各州的领导人或其他有地位的人。总之：

> 现代的邮政服务系统采用了普遍收取和寄送的理念，任何人都可以寄信给其他任何人。

当然，“普遍”只有局限性。比如，某个国家可能仅允许国内通信。这种限制的结果就是形

成了一个封闭区域，国民享受到了国内便利的通信服务，但同时，他们与国外人士的通信就不容易了。即：

> 基础设施的服务范围局限在一个封闭的区域内，在该区域的所有人都可以享有它带来的便利。

电报

电报给通信技术领域带来了巨大的改变，因为它使用电流传递信息。更重要的是，电报改变了基本的通信结构，因为它提供了高速的信息传递服务。在城市间发送电报仅仅需要几分钟，而不是数天或数周。高速的递送服务使那些通过邮件无法做到的商业活动成为可能。

> 由于电报提供了高速的信息传递服务，所以它改变了基本的通信结构。

电话

现代电话系统也给通信基础设施带来巨大的变化，它使通信更直接更即时。不像电报系统仅通过一台电报机服务整个城镇，现代电话系统在每个家庭或办公室里都有一部或几部电话，并允许两个人之间直接通信。要理解直接通信的优点，我们比较一下电话与电报。人们不用去电报局请操作员发送消息，而是直接拨打一个号码。为了明白即时访问的意义，想像一下试图通过电报进行一次对话。发送方不可能知道接收方是否能收到电报（例如，接收方是否外出），也不可能知道接收方是否需要时间来考虑如何回答，或者接收方是否收到了这则消息。如果改用电话，这两人就能直接通信。即时通信使交谈成为可能，这就好像双方在一起一样。

> 电话将通信服务扩展到单个家庭和办公室，并提供交谈所需的即时通信。因此，电话服务改变了通信结构。

11.4 Internet基础设施

Internet给通信基础设施带来了另一个重大的改变，即通用性。与以往的通信系统不一样，构建Internet的初衷并不是为了一些特定的服务或者某些特定的人群。Internet上的服务软件分为两类，一类是基础软件，它们用来实现计算机间的通信，能够为其他任何服务所用；另一类是提供高级服务的应用软件。结果如下：

> 由于基本的Internet通信设施既通用，效率也高，所以几乎所有的网络应用都可以用在Internet上。

11.5 Internet提供多样化的信息服务

当人们第一次接触Internet时，他们通常使用一种特殊的服务。有时，人们希望Internet与以往用过的任何网络都不同，但他们惊奇地发现Internet提供了在其他许多网络上也可用的服务。人们可以通过Internet给手机发送电子邮件，也可以接收手机发来的消息。此外，人们能通过Internet获取气象图、最新的笑话等多种信息，还能听音乐、看电影。

虽然Internet用户只在某一时段查看Internet的信息和服务，但Internet的设计者明白，他们不能按一些特定的服务来设计Internet技术，Internet必须支持各种各样的服务，其中大部分服务都是在建立Internet之后才出现的。简言之：

> Internet提供各种各样的服务，当前使用的大部分服务在最初设计Internet时还没有发明。

11.6 TCP/IP提供通信机制

TCP/IP软件给Internet提供了基本的通信机制，并形成了所有服务依赖的基础。TCP/IP的

灵活性是 Internet 成功的关键。因为 TCP/IP 既通用又高效，所以它能支持许多服务。又因为 TCP/IP 的灵活性和健壮性，所以它能和许多基本通信技术并用。最为重要的是，TCP/IP 的灵活性使科学家和工程师们能将 TCP/IP 和计算机、网络技术，设计 TCP/IP 时还不存在的服务一起使用。

从本质上讲，TCP/IP 提供了一套通用的通信机制，许多服务都是建立在这套机制之上的。此外，TCP/IP 的灵活性足以承受 Internet 二十年的增长，足以容纳起初并没预见到的计算机技术和网络技术，也足以支持不断变化的服务。总而言之：

> Internet 是全球的信息基础设施。虽然 Internet 提供了许多服务，但它的主要优势在于为其设计的 TCP/IP 软件，该软件能连接计算机、网络和服务的交换。

11.7　个人见解

我们居住在大学城，因此，在 20 世纪 80 年代的时候，我女儿在学校的物理实验室找到了一份暑期工作。我并不觉得惊讶。她工作后不久，在一次晚餐中，她宣布，作为她工作的一部分，她正在学习使用在 Internet 上运行的一种计算机程序。

我对她谈论 Internet 的方式感到奇怪。不象我，她并没有把 Internet 看成一个大型的计算机科学研究计划，而是看成一种服务，这就象对电话系统的看法一样。她使用 Internet，却并不知道它的历史或它的基本技术。

如今，用户们希望 Internet 无处不在，希望 Internet 易于访问，诸如电子邮件和即时通信方面的服务都已经融入了人们的日常生活，人们不再需要专家的指导就可以轻松地访问 Internet。总之，Internet 已经成为了基础设施的一部分。

第三部分　Internet 工作原理

Internet 内部揭密

Internet 基础设施的底层技术
和基本功能的诠释

第12章 包交换技术

12.1 引言

从本章起，我们将开始探索Internet使用的基本通信技术。本章将描述所有计算机网络用来传送数据的基本机制，解释这些方案成功的原因。后续章节将展示Internet如何使用这些机制。理解网络如何运行是非常重要的，因为它将解释哪些高级服务是可行的。

12.2 共享节省开支

计算机网络往往不是用一根导线将每一对通信计算机连接起来，而是多台计算机共享底层的硬件设备。共享是出于经济上的考虑：

> 多台计算机共享一条传输路径节省了开支，因为这样使用的导线和交换机数会较少。

12.3 共享导致延迟

共享传输路径并不是什么新想法，而且也不仅限于计算机网络领域。例如，一些商店要求顾客共享传送服务。商店分为两部分，即供顾客选购的大堂和存货的仓库。一条慢速移动的传送带连接这两个部分。

当有顾客提交了订单，这份订单就会被发送到仓库，在这里，店员将订购的商品发到传送带上。然后，传送带将这些商品送到大堂。这类商店通常要求店员在货物传送到下一环节之前，依次完成每份订单的发送任务。

正常情况下，共享的传送带系统会运行得非常好，因为每份订单都比较小。如果一名顾客订购一套国际象棋，而另一名顾客订购一台烤面包机，那么象棋将会先于烤面包机到达大堂。然而，当大订单出现时，传送带将延误后续的大量订单。比方说，我们想像一下，一名顾客在订购了一台烤面包机后发现传送带正忙于传送48把折叠椅，而商店又不允许传送带混合传送各订单的物品，因此，这名顾客不得不等待所有折叠椅都传送到大堂后才能收到自己的烤面包机。

要点是：

> 在一条共享路径上授予一方独占的访问权是不实际的，因为这可能延迟所有其他方。

12.4 共享线路

在计算机网络中，两台计算机间的传输路径由导线[①]构成。许多年前，设计计算机硬件和电话系统的工程师们达成了共识：一条导线一次只能传送一个数据，大量共享一条线路的设备都排队使用它。比如，在一些早期的电话系统中，一个街区的用户公用一条电话线路。如果两个邻

---

① 虽然光纤也能够作为传输介质，且传输路径也可能包含小型电子设备，但是我们仍将使用“导线”这个术语来简化这些说明。

居正在通电话，那么该街区的其他人就无法拨打电话了。类似地，当某个网络中的两台计算机传输数据时，所有其他计算机不得不等到它们之间的传送完成后才能使用该网络。

12.5 可选的通道

工程师们设计了几种解决共享资源问题的方案。有线电视系统采纳了其中一种方案。有线电视公司利用多通道技术在同一线路上同时传输多个信号。从技术上讲，该方案与第 8 章讨论的调制解调器方案类似：每一条通道使用一个频率，调制某个频率上的载波就可以编码其对应通道上的信息。有线电视公司将所有通道上的编码信号混合在一起，然后通过通信线缆传送出去。

电视接收器包括一个分离各频率输入信号的电路。无论何时，当有人把电视调至一个频道时，电视接收器就会从接收到的混合信号中提取选中频道的信号，忽略其他频道的信号。

12.6 轮流共享

虽然利用多通道技术混合在一条共享线路上的所有信号可以构建一个计算机网络，但大部分网络技术并非如此，是在传统的轮流思想上做了一些改进，即一次只允许一台计算机访问共享资源。从上述传送信的例子我们可以看出，共享规则必须从真定义，否则一台计算机长时间独占共享资源就会延误所有其他计算机。

为了避免较长时间的延迟，网络技术限制了一台计算机每次所能传输的数据量。这种思想出现在 20 世纪 60 年代，它被称作包交换技术（packet switching），一次传送的数据单元称作数据包或分组（packet）。图 12-1 说明了计算机如何使用包交换技术。

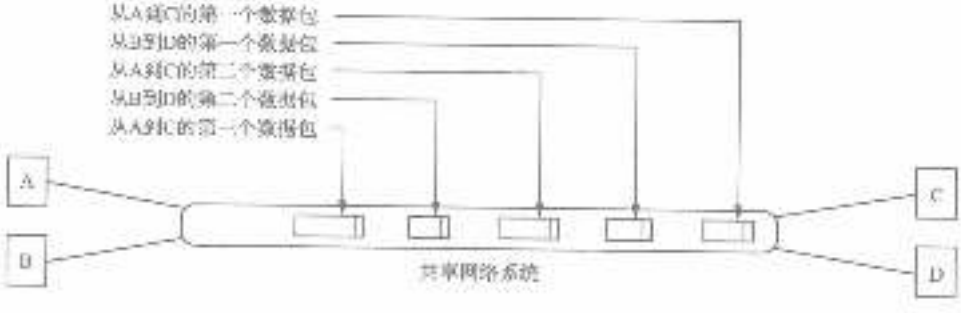


图 12-1 一队数据包通过包交换网络的示意图，其中计算机 A 与计算机 C 通信，而计算机 B 与计算机 D 通信

图 12-1 中共有 4 台计算机连入同一网络，假设计算机 A 发送数据给 C 的同时计算机 B 发送数据给 D。A 和 B 首先都需要把输出数据分解成数据包，然后轮流发送数据。首先 A 发送一个数据包给 C，然后 B 发送一个数据包给 D，B 发送完第一个数据包后，A 再发送第二个数据包，以此类推。如果 A 发送的数据量比 B 少，那么它将先于 B 结束传送，而后，B 继续发送完余下的所有数据包。

12.7 包交换技术可以避免延迟

局域网和广域网都使用包交换技术。为了理解包交换技术是如何避免延迟的，可以考虑一下上述传送带的例子。假设仓库中的店员完成订单的方式为：首先发送一份订单中的一件商品，然后再发送另一份订单中的一件商品，如此往复。如果某个顾客订购了一件小的商品，那么这件商品很快就能被放到传送带上。而如果某个顾客订购了 48 把折叠椅，那么在其他订单中各有一件商品被放到传送带上之前，这个订单中只有一把折叠椅能够被发送。

计算机支持同样的思想。计算机可以方便地将数据划分成片。如果A要向C发送一条长消息，那么A会把这条消息划分成许多数据包。如果B要向D发送一条短消息，那么这条消息只会形成一个单独的数据包或者少量的数据包。当A发送了一个数据包后，B可以发送它的数据包。因此，B不必等待A发送完所有的数据包就能有机会发送它的数据包。于是，短消息不必等待长消息全部发送完成就能进行发送。

12.8　必须标记每一个数据包

网络中传送的数据包都源于一台计算机，并以另一台计算机为目的地。计算机中的硬件监控经过的所有数据包。当它检测到某个数据包的目的地是本地计算机，它就会捕获该包，然后把该包的一份拷贝存放在计算机内存中。最后，硬件通知本地计算机接收到一个数据包。

为了使网络硬件能够区分数据包，每个数据包都必须遵循相同的格式。数据包以包头(header)开始，其后跟着数据。数据包包头标识数据包由哪台计算机来，并送到哪台计算机去。

12.9　计算机地址

网络中的每台计算机都有称作计算机地址(address)的一个惟一数字。为了便于识别一对通信计算机，数据包包头包含两个重要的地址：发送这个数据包的地址和接收这个数据包的地址。发送方地址称为源地址(source address)，而接收方地址称作目的地址(destination address)。网络硬件通常使用数字地址来发送和接收数据包。

计算机地址使用的数位取决于特定的网络技术。有的技术采用只包含几个数位的数字，而其他一些技术则使用16位的数字。这个重要的思想是：

> 计算机网络中的每台计算机都有一个惟一的称作地址的数字。数据包包含发送它的计算机的地址和接收它的计算机的地址。

12.10　数据包的大小不都是一样的

虽然包交换技术限制了一个数据包包含的数据量，但它可以传送小于最大值的任意大小的数据包。例如，一些网络应用允许用户通过敲击键盘与远程系统交互。这些应用在用户敲击键盘后立即将这次键击放在一个数据包中并发送给远程系统，产生较大数据的其他应用在传送的时候选择较大的数据包。

12.11　数据包传输似乎是即时的

在大部分包交换网络中，数据包的传送速度非常快。例如，在一个普通的局域网中，两台计算机间一秒钟能传送上千个大数据包，而传输小数据包所用的时间还要少些。对人们来说，千分之一秒内发生的事情似乎都是即时的。例如，几个人在使用接入单一、共享的网络的计算机时并不会察觉到延迟。当一个人在远程计算机上运行着一个文字处理程序时，另一个人可以访问远程数据库。每个用户在输入文字，操作鼠标时都觉得像是在本地使用计算机一样。

总之：

> 包交换系统使多对计算机能够通过一条共享通信线路来进行通信，且产生的延迟非常小。这是因为这项技术将每个会话分成许多小的数据包，并且安排在这些计算机在共享线路上轮流发送数据包。

12.12　共享是自动的

包交换技术允许计算机可以随时发送数据。一台计算机可以在其他计算机没有准备好使用

网络之前就开始发送数据。当只有一台计算机使用网络时，它就可以连续发送数据包。当有另一台计算机准备传输数据时，共享就开始了。两台计算机轮流发送数据包，公平地共享网络。如果第三台计算机准备好了，那么三台计算机公平共享网络。当一台计算机停止发送数据后，网络会自动调整，例如，如果有三台计算机平等共享网络，且其中一台完成了数据发送任务，那么剩下的两台就轮流发送数据。

更为重要的是，任何计算机都不需要知道到底有多少其他计算机在同时使用网络。关键思想是：

> 因为当一台计算机开始或结束发送数据时，包交换系统能够立即自动调整以适应新的传送情况，所以每台计算机在任何时候都很平等地共享网络资源。

12.13 网络硬件处理共享

网络接口硬件能够自动处理共享，即网络共享不需要任何"计算"，计算机在开始使用网络之前也不需要相互协同，计算机可以在任何时候生成数据包，当有数据包准备好时，计算机的接口硬件等待它的发送轮次，然后将这个数据包发送出去。因此：

> 从计算机角度来看，平等的共享网络是自动实现的，即网络接口硬件会自动处理所有的细节。

12.14 许多设备都能使用包交换技术

诸如收银机、摄像机、条形码扫描器和磁带阅读器等设备都可以接入包交换网络。许多制造商销售的打印机也能接入网络。由于一台网络打印机可以被所有联网的计算机访问，所以使用网络打印机可以降低成本：每台计算机无需一台独立的打印机。

诸如打印机这样的设备需要特殊的硬件才能接入网络，尤其是，每台接入网络的设备都需要一台小型计算机，通常这里指的是微机。这些计算机从网络上接收包含设备控制指令的数据包。总之：

> 虽然能直接连接到计算机网络的设备有很多种类型，但是每种设备都必须包含处理通信的小型计算机。

12.15 与 Internet 的关系

跟大多数计算机网络一样，Internet 采用的是包交换技术。Internet 硬件包括多用户共享的物理线路。包交换允许许多通信并发进行，即无需等待其他通信结束就可以开始某个通信。结果，无论用户何时在 Internet 上传送数据，发送方的网络软件会把数据划分成若干数据包，而接收方的网络软件必须重组这些数据包。例如，文字处理器产生的文档必须分成若干数据包在 Internet 上传输，然后这些数据包在接收方被重新组成一个完整的文档。总结如下：

> 所有在 Internet 上传送的数据都被分解成数据包，发送方把消息或文档分解成数据包，然后，把这些数据包发送到 Internet 上。接收方将接收到的数据重组成原始消息。不同计算机产生的数据包可以在 Internet 上同时传输。

12.16 小结

包交换技术是计算机网络使用的基本技术，该技术确保所有计算机能够平等地共享网络资源。在数据发送到网络中之前，它必须被分解成若干数据包。每个数据包包含一个包头，包头指定了该包应该发送到那台计算机，而目的计算机用称作计算机地址的数字表示。共享网络的计算机轮流发送数据包，在每一轮中，一台计算机发送一个数据包。虽然诸如打印机之类的设备可以接入网络，但这样的设备必须包含控制网络通信的微型计算机。

第 13 章　Internet：网中网

13.1　引言

第 12 章介绍了包交换技术，解释了为什么在一条共享传输路径上，将长信息分解成短小的数据包能降低计算机的传输延迟。本章将介绍多个包交换网络是如何互联成 Internet 的，而 Internet 运行起来就像一个单张巨大网络。

13.2　网络技术互不兼容

为了满足速度、距离和造价等方面的不同需要，现存有许多种包交换技术。造价低的网络的运行速度通常都低于高成本网络的运行速度。设计者们并不打算让所有的设计都兼容，即是说，诸如电压、表示计算机地址的数字等细节通常都是不同的。因此，仅仅连接两个或更多较小网络的线路是不可能组成一个大型网络的。

为了理解技术上的不兼容造成的后果，考虑一家拥有两个互不兼容的局域网的大型企业。比如，我们假设财务部门和生产部门分别使用一种类型的网络。当财务部门需要一些存放在生产部门计算机中的信息时，首先必须把这些信息复制到外部存储设备中（如闪存），然后再人工送到财务部门。如果我们能把两个网络互联起来，且两个网络的计算机之间可以以电子的速度传送信息，而无需人工传送闪存，那么互联起来的网络将节省大量的时间和金钱。

13.3　解决不兼容问题

解决网络不兼容问题有两种可能的方案：

- 选择一种满足整个公司需求的网络技术，公司上下都使用这一选定的技术。
- 允许各部门选择最适合本部门需求和预算的网络技术，然后找到一种将所有类型的网络互联起来的方法。

第一个解决方案存在几点不足。首先，找到一种能满足所有需求的单一网络几乎是不可能的，即使这样的技术存在，其造价也可能会很高，这是用户难以接受的。更加重要的是，当公司成立一个新部门时，整个需求一般会发生改变。如果公司中的每个部门必须使用相同类型的网络，那么通过改变整个技术来容纳一个新成立的部门将意味着要更换掉所有现存的设备。因此，我们得出的结论是：

> 要求一家公司的所有计算机都使用相同的网络技术是不实际的，从经济的角度来也是不可行的。

Internet 采用第二种方案，它允许每个部门选择最符合自身需要的网络技术。为了包容各种各样的网络技术，Internet 提供了一种连接任意网络的机制，还提供了一种通过连接线路传输数据的软件。

13.4　两个基本概念

两种简单的方法将有助于解释 Internet 中使用的一些网络互联技术。第一种方法将说明解决

远程连接问题的可能性，第二种方法将解释连接与不兼容的网络的可能性。

网络连接可以扩展

连接局域网的每台计算机都需要一条从计算机到局域网的线缆。虽然计算机与局域网间的线缆通常都很短（如几英尺或办公楼里的走廊的宽度），但它是可以延伸的。图 13-1 说明了计算机连接远程网络的三种方式。

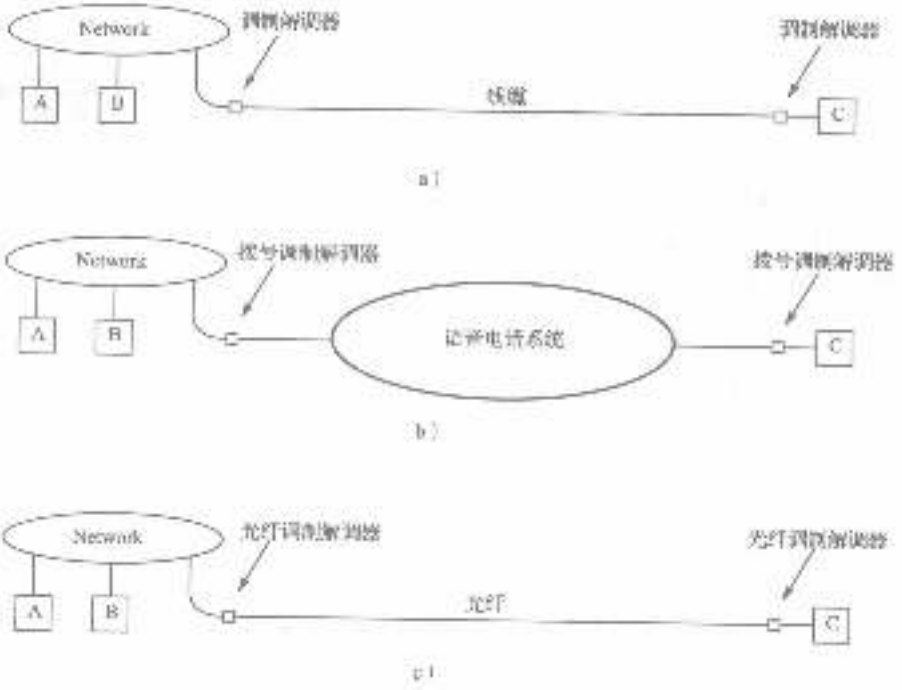


图 13-1 延伸计算机与网络之间连接的一种方法：a）使用两个调制解调器的连接线路；b）使用两个调制解调器和语音电话系统的连接线路；c）使用光纤调制解调器和光纤连线

第 6 章讨论了调制解调器是如何通过导线或电话系统传输数据的，而许多现代通信系统使用柔软的玻璃纤维作为远距离连接的传输介质。这种技术称作光纤（fiber optic）连接技术，它需要利用光而不是电子传输数据的光纤调制解调器。在发送方，光纤调制解调器通过激光器将数字转换成光脉冲，然后将其发送到光纤中。在接收端，光纤调制解调器感应光脉冲，然后把它们还原成计算机数据。

光纤连接方式具有远程传送的优势（如校园内或城镇间）。光纤有三项很有意思的性质，这些性质说明了光纤为什么如此有用。首先，光纤比较柔软，它可以像其他线缆那样安装。其次，光纤由玻璃制成，因此可以传输光。最后，由于光纤中传输的光只能在其内部反射，所以光的强度不会随传输而减弱，即是说，几乎没有任何光束能从光纤的边缘逃逸出来。

一台计算机可以拥有多条连接

13.3 节所述的第二种方法有助于直观解释 Internet 技术：一台计算机可以接入两个或两个以上的网络。第 7 章解释了包含一块网络接口卡（NIC）的计算机，这个 NIC 将计算机接入到某一网络。实际上，一台计算机可以拥有多块 NIC，每块 NIC 将计算机接入到一个网络上。所有的网络并不需要使用相同的技术。一台计算机连接到多个网络之后，它就可以发送或接收任何一个网络的数据包。

13.5　利用计算机互联网络

Internet使用专用计算机进行网络互联，图13-2解释了这一概念。用于连接网络的计算机类似于传统计算机，它也包含中央处理器、存储器和网络接口，它也具有不同的规模和速度。最小的、功能最少的模块仅花费不到一百美元，而功能强大的模块能互联多个网络，其花费高达十万美元。

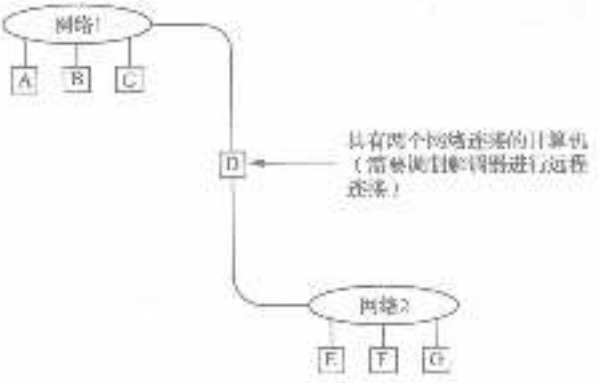


图13-2　计算机D连接了两个网络，这两个网络可以使用同一种技术，也可以使用不同的技术

尽管连接网络的计算机使用传统硬件，但是它们并不使用传统软件。由于这些计算机专门处理网络互联，所以它们不装载应用程序。比如，这些计算机不运行文字处理程序，相反，它们只有一个特定用途的软件，该软件执行与网络连接相关的任务。实际上，这些互联网络的计算机一般都被设置成一旦接通电源后就自动运行所需软件的模式。因此，当电源故障恢复后，它们可以自动重启，正常情况下，不需要任何人工干预。

总之：

> 互联网络的计算机专门用于网络互联任务，尽管它们使用传统硬件，但它们使用开机后可以自动运行的特定软件，并能一直运行下去。

13.6　互联的计算机传递数据包

互联网络的计算机执行的一个主要任务是：将数据包从一个网络转发到另一个网络。下一章将介绍数据包是如何转发的，其思想很简单：计算机从一个网络接收传给它的数据包，然后通过另一个网络将数据包传送到目的地。如图13-2所示，计算机A和G没有接入到同一个网络。如果计算机A生成一个发往G的数据包，A首先将该包发给连接这两个网络的D，之后，D从网络1接收该包，并通过网络2将包发给G。

13.7　连接网络的计算机称为路由器

连接网络的计算机上的软件需要知道每台计算机连接在哪个网络中，这样才能确定向哪里转发数据包。在只有两个网络的情况下，确定转发路径非常简单：当来自一个网络的数据包到达时，它应该是发往另一个网络。然而，当有三个网络互联时，情况就变得比较复杂。当一个数据包从一个网络传来时，软件必须从其他两个网络中选择一个网络。

选择一个网络来传送数据包的过程称为路由（routing），而连接网络并执行转发任务的计算机称作路由器（router）。下一章将讨论Internet中路由的细节。

13.8　路由器是构建Internet大厦的基石

尽管很多人都认为Internet是一个单一的、巨大的、连接许多计算机的网络，但事实并非如此。事实上，Internet由成千上万的计算机网络组成，这些网络通过路由器互联起来，每台计算机连接到上述网络之一。当一个网络中的计算机与另一个网络中的计算机通信时，它们通过路

由器传递数据包。总结如下：

> Internet 不是一个单一的计算机网络，而是由成千上万的计算机网络互联而成，连接这些网络的专用计算机称作路由器。

路由器在 Internet 中有多种用途，例如，路由器可以连接一栋建筑或至一个房间里的两个局域网；通过一对调制解调器，路由器可以互联两栋建筑或者整个校园里的局域网。由于路由器能互联多种不同的网络，所以它可以用来连接局域网和广域网。

13.9 Internet 包含多种网络

因为路由器能够互联使用不同硬件技术的网络，所以路由器结构允许 Internet 包含多种类型的网络。每个小型团体都可以选择最符合自己需求和预算的局域网技术，然后通过一台路由器就可以将自己的网络接入 Internet。大机构可以自己购买的路由器组建私有互联网。通过添加一台路由器，这个组织的私有网络就可以接入全球 Internet。

有两个原因能解释为什么连接多种类型的网络很重要。首先，由于 Internet 连接的众多机构具有不同的规模、需求和预算，所以它们采用的网络技术也不同。其次，更为重要的是，由于计算机网络一直都是研究的热点领域，所以网络技术不断发生变化。目前在 Internet 中使用的许多硬件技术在 Internet 诞生之初尚不存在，现用的许多技术在将来也会被其他新技术取代。如果 Internet 不能容纳各种类型的网络，那么它就不可能长久存在。

13.10 路由器可以互联广域网和局域网

路由器解释了 Internet 可以连接局域网和广域网的原因。例如，一个属于主要 ISP 的大型广域网提供能有效连接大量站点的远距离连接技术。术语主干网（backbone network）通常用于描述连接其他网络的主要广域网。并非所有站点都要连接到主干网，只有那些称作主干站点的节点才连接到主干网。图 13.3 说明了这种体系结构。

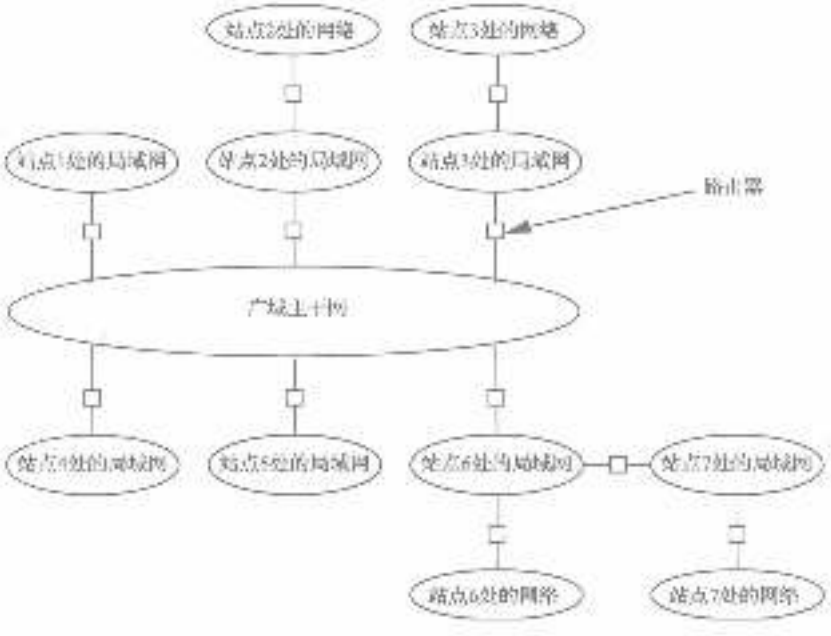


图 13.3 广域主干网示意图。在每个主干站点，都有一台路由器将局域网连接到主干网。在某些站点，额外的路由器连接附加的网络。站点 7 处的网络没有直接接入主干网

如图13-3所示，位于每个主干站点的路由器将广域主干网和局域网连接起来。站点间的路由器提供了一条连接不同局域网内计算机的路径。每个站点都可以通过附加的路由器连接到本地站点或其他站点处的附加网络。图13-3中，站点2有两台路由器和两个网络。站点6处有3台路由器：一台连接主干网，一台连接另一个网络，第三台连接站点7处的局域网。

通常来讲，将一个网络接入到Internet的最经济的方法是找到最近的Internet站点，并通过一台路由器连接到Internet。但是，并不是所有站点都有权提供这种连接。此外，跨国界的连接（例如，在位于两个国家网络间建立连接）也是不可能的。

13.11　Internet的层次结构

图13-3给出了Internet的简单结构图，实际上，即便是只包含几个网络的中等规模ISP也是通过路由器连接起来的，而大型ISP具有更加复杂的结构。除了负责运行主干网，大型ISP还负责一些覆盖较大地理范围的区域网络（regional network）和一些提供公司或住宅区接入服务的接入网络（access network）。因此，我们提出了一个问题：Internet的机构是如何组织的？

为了有助于理解Internet的结构，同时也不至于过于细节化，网络专家绘制了一种图表，其中云状物表示一组网络，例如，图13-4解释了ISP的层次结构，其中主ISP在顶层，地区ISP在中间，而本地ISP离用户最近。

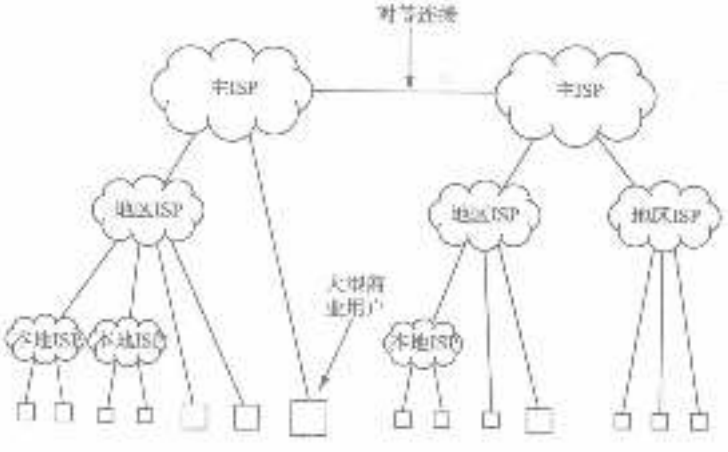


图13-4　Internet结构示意图，其中每个ISP都用云状物表示。图底部的方框表示用户，其中，小方框表示住宅区或小型商业用户，大方框代表大型商业用户

主ISP，即操作主干网络的ISP，称为一级供应商，地区ISP称为二级供应商，而本地ISP则称为三级供应商。因此，诸如AT&T或Sprint之类的供应商都运行着大型主干网络，它们属于一级供应商。

如图13-4所示，一级供应商之间的连接称作对等连接（peering arrangrment）。这两个供应商都是对等节点，因为它们各自都拥有差不多相同数量和类型的用户。对等只是就世界上最大的ISP而言，图13-4中，小型本地ISP是中等规模的地区ISP的用户，而地区ISP是一级供应商的用户。

如图13-4所示，一些居民区用户直接接入地区ISP，而其他的则接入到本地ISP。总之，本地ISP无法向大型企业用户提供服务，因为本地ISP和地区ISP之间的连接没有足够的带宽来处理大规模的通信流量。因此，大型企业用户通常选择直接接入地区ISP或一级ISP。

13.12 数据包传送的过程

为了理解 Internet 的层次结构，我们考虑地理位置相距很远的两个用户。假设每个用户都接入到本地 ISP。那么当他们通信的时候会发生什么呢？当一则消息发送出去后，数据包从发送方的计算机出发，经由接入网络到达本地 ISP。本地 ISP 将数据包转发给地区 ISP，然后地区 ISP 依次将包转发到一级供应商。如果目的主机在同一个一级供应商下，那么数据包将通过主干网到达目的主机所在地区 ISP。如果目的主机在其他一级供应商下，那么数据包需要通过对等连接传送到另一个一级主干网。数据包一旦到达目的主机所在的地区 ISP，地区 ISP 就会把这个数据包转发到正确的本地 ISP，然后本地 ISP 再将数据包投递到目的主机接入的接入网络。当然，沿途的每个 ISP 都包含多个网络和路由器，因此数据包可能会经过许多网络。

要点如下：

> Internet 是网中网，它具有一定的层次结构。一级 ISP 在顶层，地区 ISP 在中层，而本地 ISP 位于最下层。

13.13 网络互联具有革命性

使用路由器互联两个网络的方法看起来不怎么样，但它的确不错。Internet 技术出现之前，一个公司需要选择一种网络技术连接所有计算机，或者需要多个独立的网络。路由器使公司的各个部门可以选择最适合自己的网络技术，同时，路由器提供了互联这些网络的方法。

13.14 小结

虽然对于用户来说，Internet 就是一个巨大的单一网络，但事实上，Internet 是由成千上万个计算机网络通过称作路由器的专用设备互联起来的。由于路由器能够互联采用不同技术的网络，所以，它可以连接局域网和局域网、局域网和广域网、广域网和广域网。由于 Internet 是由通过路由器互联的多个网络组成，所以许多人把它称作网中网（network of networks）。

Internet 中的网络按层次结构组织。一级 ISP 形成了 Internet 的第一级，地区 ISP 形成第二级，而本地 ISP 形成了最下一级。在层次结构的每一级中，低层 ISP 是高层 ISP 的用户，当两个一级 ISP 之间要传输数据时，它们之间的关系是对等的。

通过路由器互联网络是具有革命性意义的。它允许各种不同类型的网络之间实现互联，也允许一个组织中的各个部门选择最符合自身需求和预算的网络技术。

第 14 章　ISP：宽带与无线接入技术

14.1　引言

第 13 章把 Internet 称作网中网，然而对于大部分人来说，主要关心的问题不是如何构建 Internet，而是计算机如何接入 Internet 以及如何使用它。本章描述了应用于 Internet 连接的各种技术，解释了各种接入技术提供的服务。

14.2　ISP 及其资费

我们把提供 Internet 访问服务的公司称作 Internet 服务供应商（Internet Service Provider，ISP）。跟其他公司一样，ISP 要对此项服务收取服务费，通常，ISP 收取两种费用：

- Internet 使用费
- Internet 接入初装费

在美国，每隔一段固定的时间，Internet 使用费的账单就会开给用户，即是说，ISP 在每个月至固定时间都会向用户收取一次使用费，此项费用与用户使用 Internet 服务的具体时长，跟目的主机通信的距离，以及传输的数据量无关。ISP 对已付费的用户提供报文转发服务——将来自用户计算机的数据转发给 Internet 中的目的计算机，也将目的地址为用户计算机的来自 Internet 中其他计算机的数据转发给用户计算机。

尽管使用费的收取方式规定不变，但是 ISP 将用户划分成若干等级，不同等级用户收取不同的费用。例如，相对于个人用户，ISP 通常对企业用户收取更多的费用，因为企业往往十分频繁地使用 Internet，而个人用户倾向于离散地使用 Internet。除此以外，ISP 还根据用户的 Internet 初装类型收取不等费用——具有较大带宽的用户比小带宽用户付费更多。

另一种费用就是初装费。用户与 ISP 之间通过一条专线连接起来，下一小节将讨论和解释应用于此连接的几种技术。

14.3　用户连接形成“最后一英里问题”

虽然 Internet 由各种互联的网络组成，但是在 ISP 及其用户之间直接连接是比较困难的，其原因有三。第一，因为一个 ISP 拥有许多用户，连接技术必须能够覆盖大量的物理区域。第二，由于大部分用户属于个人或家庭用户，所以采用的连接技术造价必须低廉。第三，如果用户获取信息的等待时间太长，会引起用户不满，这些技术必须能够提供较高的数据传输率。因此，工程师们处于既要提供优越的服务性能又要造价低廉的两难境界之间，这个问题在业界被称作最后一英里问题（last mile problem），而这些技术则被叫作接入技术（access technology）。

20 世纪 90 年代中期以前，仅仅只有一种技术用于把专线接入 Internet：从公共承运商租用的数字电路。从本质上讲，电话公司通过出租电话线缆中尚未用部分形成租赁线路，由于租赁线路十分昂贵，所以往往是企业级用户使用此项技术——在企业与 ISP 之间租借一条线路。一位用户马先需要支付初装费，然后按月支付线路使用费。一些 ISP 要求用户直接与承运商签订租用协议（例如，电话公司），而其他 ISP 与用户签订使用协议。无论哪种情况，用户必须按月向 ISP

或公共事业单位支付一定的线路使用费。

14.4 租用专线费用昂贵

租用专线的主要优势在于其可用性及易控制性。一条线路能够保证一天24小时，一周7天都处于可用状态。因此，如果一家企业要让它的Internet站点一直在线，那么它会选择租用一条专线。更为重要的是，由于各个线路专属于每个用户，所以线路上仅有的流量是属于租用该条线路的用户的。因此，为了确保传输的信息不受其他信息干扰，一家企业可以选择租用专线，这样有且仅有该用户的信息在此专线上传输。

租用专线的主要劣势在于其高代价，因为租用一条专线是非常昂贵的。例如，笔者所在大学租用的专线中一条的费用就高达一个月几千美元。因此，所有个人用户能够承担起这么一笔费用。

专线每月的花费依据的是线路的物理长度（例如，用户至ISP的物理距离）及其容量。由于企业使用的专线的高容量，我们把它分类成宽带（broadband）。其中的一类被称作T1线路[3]——电话公司标准的名称，T1线路指定了线路的电子规范以及位传输率。

14.5 拨号访问线路便宜但缓慢

如果一个人无法担负起租用专线的费用，那么他/她又该如何访问Internet呢？幸好，这里有几个可以选择的对象，其中，最便宜的Internet接入技术由传统的电话系统组成。大部分ISP提供基于电话的服务，这种服务被称作拨号访问服务（dial-up access）。

要使用拨号访问，计算机必须有一台调制解调器连接到电话系统，就像普通的电话那样。除此之外，计算机还必须安装控制调制解调器的软件。当软件运行时，它命令调制解调器按ISP提供的电话号码进行拨号。在ISP方，另一个调制解调器响应该拨号并同意其接收Internet数据报。图14-1示意了这种方法。

图14-1 通过拨号调制解调器和电话网络访问Internet。拨号访问服务没有持续Internet连接服务费用高

虽然拨号连接提供了Internet上的所有服务，但是它的响应时间比较缓慢。概括成一句话就是：

> 通过传统电话系统接入Internet是一项费用最少的接入技术，但是它的访问速度是比较慢的。

14.6 宽带提供高速连接

用户都想要数据被快速递送，毕竟，较快的传输意味着较少的等待时间。拨号与专线访问服

---

③ 笔者所在的城镇中有一台复式公寓这样做广告，“每个房间都有T1线路”。

务的速度有着巨大的差异，在传输图像的时候，它们之间的差异是非常明显的。例如，有人要发送给笔者一张高清晰照片，其大小为3 525 353字节，那么在传输率为33.3 Kbps的拨号连接上传送需要将近14分钟，而在T1线路上传送的时间仅为18秒。

我们用术语"宽带"表示一项使数据高速传输的技术。例如，T1线路被归作宽带是因为它每秒能传送超过一百万比特；而拨号线路不能称作宽带的原因在于其传输速率被限制在56 Kbps以内。宽带提供的高速允许用户在没有较长延迟的情况下浏览更多的数据。因此，采用宽带接入技术的用户比采用拨号接入技术的用户更加可能去等待一整张图像的下载。

14.7　重要概念：持续连接

从最广义的角度讲，拨号接入技术与专线接入技术似乎提供完全相同的服务。一旦建立起拨号连接，用户就可以发送和接收数据包，也可以与Internet中的任何目的主机进行通信。尽管如此，它们之间还是存在两个基本的不同之处。专线接入技术提供：

- 高速数据传送
- 即时访问与持续可用性

虽然不同的用户对"高速数据传送"这一概念的理解各异，但是专用线路较拨号访问提供了一个更为重要的改进：即时访问与持续可用性。这个改进到底起到了什么作用呢？我们不妨考虑一位拨号用户在访问Internet时遵循的步骤。在该用户开始访问Internet前，他必须等到电话处于空闲状态。在大多数情况下，该用户还必须首先打开计算机，等待计算机启动后，使用调制解调器拨号ISP的软件。在ISP方，一个调制解调器对用户的来电做出应答，然后便开始与用户的调制解调器进行通信。之后，用户还必须等候通信双方的调制解调器交换测试消息，以此校验电话连接的质量以及协商双方的通信速率。一旦两个调制解调器取得了同步，其他Internet软件就可以开始发送数据包了。经过这些步骤的等待后，用户终于可以运行访问Internet的应用程序了，习惯上，该用户会启动一个浏览器，当然这也是需要一些时间。

现在再看专线接入技术与拨号接入技术有那些大的差异。由于专用连接线路一直被置于某个合适的地方，所以用户不需要等候调制解调器拨号、ISP方调制解调器应答、双方调制解调器同步、或者其他Internet软件开始运行。取而代之的是，计算机保持对Internet的持续连接，并且一直准备着发送或接收数据包。因此，用户根本不需要等待连接Internet。为了更快地访问Internet，采用专线接入技术的用户常常让他们的电脑一直开着，同时使浏览器在后台一直做好连接准备。因此，计算机可以持续访问Internet——所有的延迟都可以被忽略掉，而用户需要的仅仅是按动按键或敲击鼠标，然后就可以访问Internet了。

14.8　即时访问改变Internet使用方式

即时访问与持续可用性影响了人们使用Internet的方式，这个影响是比较有趣的。用户不是把访问Internet看作是一个耗费时间的活动，而是把它看成一种非常容易的事。采用拨号接入技术的用户往往计划在一天的某个特定时间使用Internet，而采用专线接入技术的用户可以非常容易地调用Internet服务，而无论何时需要这项服务。因此，使用专线的用户不用在单一会话上安排所有Internet使用时间，结果，这些用户在Internet的使用上就变得闲散了——无论何时有需要，他们都可以用Internet一次完成一个任务。例如，专线接入技术可以非常轻易地帮助人们浏览当前天气状况，类似地，专线访问技术使查收电子邮件非常容易——在屏幕上的图标可以告诉用户电子邮件何时可用，以及一次简单的点击鼠标就足以阅读和显示消息。结果如下：

一条专用的Internet接入线路在提供每天24小时的即时访问服务的同时，改

变了人们使用 Internet 的方式，因为它鼓励人们频繁及频繁的使用。

14.9 现代技术提供便宜的专线接入服务

ISP 与工程师们意识到客户都想要高速数据传输以及专用 Internet 连接服务。因此，在 20 世纪 90 年代，工程师们都在努力地寻找这样一种接入技术，该技术在具有高速、专用连接技术的优点的同时比租用专线具有更低的费用开销。有三项技术先后出现，现在，它们正被 ISP 广泛使用：

- 电缆调制解调器（Cable Modem）技术——与有线电视使用相同配线
- 数字用户线（Digital Subscriber Line，DSL）技术——与传统电话使用相同配线
- 无线接入（Wireless access）技术——使用无线电信号传送数据

电缆调制解调器

在 Internet 普及前，有线电视已经布线到大部分家庭中。尽管同轴电缆设计的初衷是类似模拟视频信号而非数字数据，而且信息在同轴电缆中是单向传播的（朝着客户方向），但是该系统现在已经发生改变。一项被称作电缆调制解调器的新技术油然而生，它使用电缆布线提供 Internet 服务。电缆调制解调器就是一个把用户计算机与线缆系统连接起来的小电子盒。除了客户方需要电缆调制解调器外，电缆公司方也同样需要。电缆调制解调器编码待传数据，然后在同轴电缆上进行传送。电缆调制解调器的作用是避免数据传输不会干扰电视信号，在有线电视信号传送的同时使数据传输成为可能。

电缆调制解调器相比其他技术有四方面的优势。第一，它们传送数据比拨号连接方式快得多。第二，它们提供连续、即时的连接服务。第三，它们不像租用数据专线那样昂贵。第四，它们用到了业已存在的布线系统。

电缆调制解调器技术的主要缺点在于它们要求大量用户"共享"同一线路。电缆公司把近百人的用户组（例如，来自于 100 个居民的调制解调器）连接到同一底层线电缆系统，然后对这些调制解调器进行排序并轮流发送数据。在某个时间段里，如果组内仅有一个用户使用计算机上网，那么这个用户将占据全部带宽，数据将以最高的可能速度进行传输。然而，如果所有用户都同时使用各自的调制解调器，那么调制解调器将轮番发送数据，这也就意味着数据传输速率要降下来。即便传输速率降到最低，电缆调制解调器仍然比拨号方式快得多。但是当许多用户同时访问 Internet 时，有效数据传输率一定会发生显著的变化。因此，一个用户享受 Internet 服务的质量依赖于其他用户（例如，在邻近区域内）使用 Internet 的程度。

DSL 技术与电话

电话也已经布线到大部分家庭中，电话公司经过调查后发现已有的布线已经可以用来传输数字数据了。起初，电话系统是用来传送模拟声音信号，现在，这个系统已经被扩展了。一项被称作数字用户线（Digital Subscriber Line，DSL）[①]的技术浮出水面，它提供基于电话线路的高速 Internet 访问。

我们似乎已经讨论了拨号电话系统是如何用作 Internet 访问的，但是 DSL 技术根本没有用到电话系统，相反，它仅仅使用了电话系统的底层布线。

为了更好地理解什么是 DSL，你必须知道一条双绞线就能把各个居民区（或者企业）与电信局连接起来。DSL 技术并没有使用拨号调制解调器对数据进行编码，而是通过在电话线上发

① 存在一些 DSL 技术的变体，例如居民用户往往使用 DSL 技术的一种特殊类型，称作非对称数字用户线（Asymmetric Digital Subscriber Line，ADSL）技术。

送电信号来高速传送数据。这样的设计是非常精妙的，因为我们可以仔细选择 DSL 使用的电信号，以致它们不会干扰电话信号。安装的过滤器（filter）用来确保不会出现偶然的干扰。所以，除了传送标准的电话信号外，在一对电线上传送数据完全是可能的。

从某种意义上讲，DSL 信号是不可见的，因为，在任何时候，它们在电话线上传输的同时都没有影响到正常的电话服务。因此，当有人使用 Internet 时，电话仍然可以响铃，也可以接听。同样地，人们也可以拨打电话，这并不影响 Internet 的使用。总结起来表达如下：

> 虽然 DSL 技术通过传统电话服务使用的电话线发送数字数据，但是它却不产生任何干扰。打电话不影响使用 Internet；使用 Internet 也不影响打电话。

除了允许同时使用 Internet 与电话，DSL 技术还具有其他几个优点。首先，DSL 提供高速数据传输，这与电缆调制解调器技术的可用性相似。更为重要的是，DSL 用户之间不用共享同一带宽——用户与电话公司之间传输的数据是相互独立的，不会互相影响。其次，与电缆调制解调器技术相似，DSL 提供连续、即时的连接服务。再次，DSL 较租用专用线路更为便宜。最后，DSL 使用的是已经存在的电话线路系统。

14.10　无线访问的可用性

虽然 DSL 与电缆调制解调器技术似乎完全解决了最后一英里问题，但是它们都具有一定的局限性。DSL 存在距离限制，即传输的信号不能像电话信号传送得那么远。因此，DSL 无法覆盖到郊区。同样，一些家庭并没有安装有线电视。结果，基于经济因素，电缆和电话公司可能无法对所有区域提供 Internet 接入服务。

为了覆盖到远程区域，工程师开发出了无线接入技术（wireless access technology）。尽管所有无线网络都通过无线电波承载数据，但是目前已经发展出了多种技术。其中一些无线网络采用点对点技术，即是说，在两个通信站之间（例如，像餐的住所与 ISP 之间）使用特殊的天线直线发送信号；而另一些则是采用广播方式传送数据（例如，在一台发射机和相邻区域的住房之间）；其他无线网络在 ISP 与处于任意位置的用户之间通过绕地卫星传递数据。要点是：

> 多种无线接入技术已经开发出来，以满足各方面的需要。

14.11　Wi-Fi 与 3G 无线技术

目前有两种无线技术已经变得非常受欢迎：

- Wi-Fi 局域网
- 3G 无线蜂窝网络

Wi-Fi 网络提供小范围内（例如，在一间房屋里或在一家商店里）计算机间的连接服务。为了提供 Wi-Fi Internet 接入服务，人们必须购买并且安装一种被称作 Wi-Fi 接入点（Wi-Fi access point）的设备，接入点将无线网络连接到传统的有线 Internet。从本质上讲，接入点扮演的是路由器的角色，它在无线与有线网络之间传递数据包。

现在，大部分笔记本电脑都内置了一个 Wi-Fi 网络接口卡，当笔记本电脑距离接入点足够近时，此卡允许笔记本电脑接入 Internet。像咖啡店这样的行业把提供免费的 Wi-Fi 接入服务作为吸引顾客的一种方式，有了这种服务，顾客在享用咖啡的同时就可以用笔记本电脑访问 Internet。其他场合（如机场）也提供 Wi-Fi 接入服务，但需要用户支付一定的服务费。

遍布世界各地的电话公司已经对 Internet 服务的强大需求做出了相应的回应，它们设计了各自的数据传输方案，我们把这些方案统统称作第 3 代无线技术（3rd Generation wireless）或者 3G 无线技术（3G wireless）。3G 电话配备有能够显示文本或图像的显示屏，而且电话公司已经

计划提供更多的数据服务。尽管 3G 无线技术构建在蜂窝电话技术而并非 Internet 技术之上，但是 3G 服务与 Internet 是相互连接的。因此，把电子邮件发送给 3G 手机或者接收来自 3G 手机的短消息并且通过 Internet 把它发送出去，这些情况都是完全可能的。

14.12 个人见解

身为网络研究员，我一直有机会在早于各种新技术商业化之前就尝试它们。例如，在 20 世纪九十年代末，我参与了一项实验，该实验使用 DSL 技术以每秒六百万比特的速率向我的住所发送数据。目前，我的实验室与住所之间有两个专用连接线路，各自的传输速率都达到了 2 240 000 比特每秒。不管怎么说，从 20 世纪七十年代末我的住所可以全天二十四小时享受 Internet 服务。

几年前，我参加了由一位高技术产业领域的名人主讲的讲座。尽管他的讲座的公告上说，他将讨论 Internet 未来的发展前景，但是讲座的结果是令人失望的，因为这个讲座中的重点是那些已经派上用场的技术。到讲座快结束的时候，显然，我享受了许多年的这些设备对这位演讲人来说似乎是一样新的事物，因为他告诉听众，当他从缓慢、笨拙的拨号接入技术转换成持续的宽带连接时，他对 Internet 的认识才发生改变。由于宽带接入已经成为标准技术，而且 Internet 服务就像电话服务那样平常，所以人们很难想像离开 Internet 之后如何生存。

第 15 章　IP：建立虚拟网络的软件

15.1　引言

第 13 章讲述了 Internet 作为网中网，是由路由器这种具有特殊用途的计算机将网络连接起来而成的。当然，仅仅将硬件连接起来并不能构成 Internet。互相连接的计算机需要软件才能进行通信。本章将描述该硬件作为一个巨大的单一网络运行的基本软件。

15.2　协议：通信协定

两个人如果不说同一种语言是几乎无法沟通的。同样地，两台计算机如果不使用同一种语言也无法通信。通信协议（communication protocol）是用来规定两台计算机用于交换信息的共同语言的协定。这个术语来自外交词汇，其中协议是指外交活动所要遵守的规则。

计算机通信协议精确地定义了通信事宜。例如，协议规定计算机所能发送的消息的确切格式和意义。协议还规定了计算机何时发送消息，以及收到一条消息时应当如何应答。

15.3　基本的功能协议：网际协议

Internet 中使用的最关键的通信协议叫做网际协议（Internet Protocol，IP）。网际协议详细地定义了计算机如何进行通信的细节。它精确地规定了如何生成一个数据包，以及路由器如何向目的地转发各个数据包。

每个连接至 Internet 的计算机都必须遵守网际协议的规定。计算机生成的数据包必须使用 IP 协议规定的格式。当一台计算机收到一个数据包，这个数据包一定是发送的原始数据包的一个精确副本，并且仍然符合 IP 协议格式。更进一步来讲，Internet 中的每一个路由器都希望送给它的数据包符合 IP 协议格式。

15.4　每台机器上的 IP 协议软件

计算机硬件并不理解 IP 协议。因此，一台计算机连接到 Internet 上并不意味着它能够使用 Internet 服务。计算机需要 IP 协议软件才能在 Internet 上进行通信。实际上，每一台想要使用 Internet 的计算机必须运行 IP 协议软件。

IP 协议是基础：所有的 Internet 服务都使用 IP 协议来发送和接收数据包。因此，每台计算机通常都包含一个所有应用程序共享的 IP 协议软件的副本。在计算机中，操作系统总在内存中保存 IP 协议软件的一个副本，以便随时发送和接收数据包。总之：

> 因为所有的 Internet 服务都使用 IP 协议，计算机在访问 Internet 之前必须先安装 IP 协议软件。

15.5　Internet 数据包称为数据报

为了区别 Internet 数据包和其他网络的数据包，我们称符合 IP 协议规定的数据包为 IP 数据报（IP datagram）。之所以取这个名字，是为了使人们对 Internet 数据包传输服务如何处理数据

包有一个直观的认识。顾名思义，Internet 处理数据报的方式与电报局处理电报的方式相同。当发送端计算机生成一个数据报并把它放在 Internet 上进行传送时，发送端可以自由地继续去做其他事情，就好比一个人交给报务员一个消息去发送之后可以自由离开电报局一样。数据报在 Internet 上的传输是与发送方独立的，类似于报务员发送一个电报至目的地与发送电报的人无关。总之：

> Internet 上传送的数据包必须符合 IP 协议规定的格式。这样的数据包被称为 IP 数据报。

15.6 巨型网络的设想

尽管 IP 协议定义了许多通信细节，但是它有一个重要的目标。每一台 Internet 上的计算机在安装了 IP 协议软件之后，都能够生成 IP 数据报并且将它发送给其他任何一台计算机。从本质上说，IP 协议将一组网络和路由器组合成一个无缝的通信系统，使得 Internet 像一个巨大的单一网络一般运行。图 15-1 说明了这种思想。

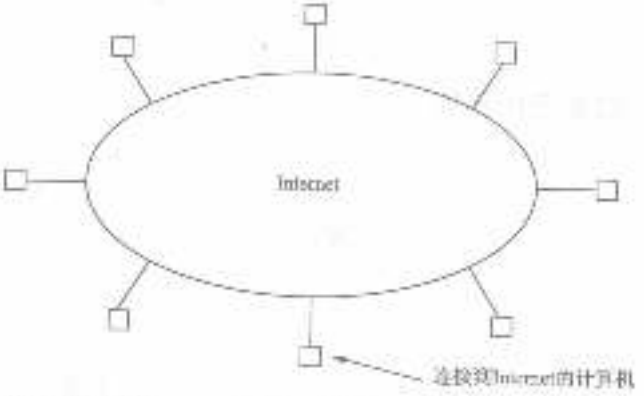


图 15-1 IP 协议软件提供的 Internet 视图。用户和应用程序认为 Internet 是一个允许任意数量的计算机进行通信的巨大的单一网络

计算机科学家使用术语虚拟（virtual）来描述这种超出硬件所能提供的，表现出更强大的计算能力的技术。Internet 是一个虚拟网络（virtual network），因为它呈现为一个巨大的单一网络。尽管 Internet 自身是一个网中网，但是 IP 协议软件处理了细节问题使得用户认为 Internet 是一个单一的网络。用户始终察觉不到 Internet 的网络和路由器，正如电话用户始终不清楚构成电话系统的线路和交换机一样。

要点是：

> Internet 运行起来像一个连接了几百万台计算机的单一的网络。IP 协议软件允许任意一台计算机向其他任何计算机发送 IP 数据报。

15.7 Internet 的内部结构

尽管用户认为 Internet 是一个巨大的单一网络，但是 Internet 包含着用户并不可见的复杂的内部物理结构。图 15-2 说明了这个概念。

当某个数据报在 Internet 上由一台计算机传输至另一台计算机时，它必须通过一条物理路径。在这条路径的每一步，该数据报要么在一个物理网络中传输，要么通过一个路由器到达另一个网络。最终该数据报到达它的目的地。

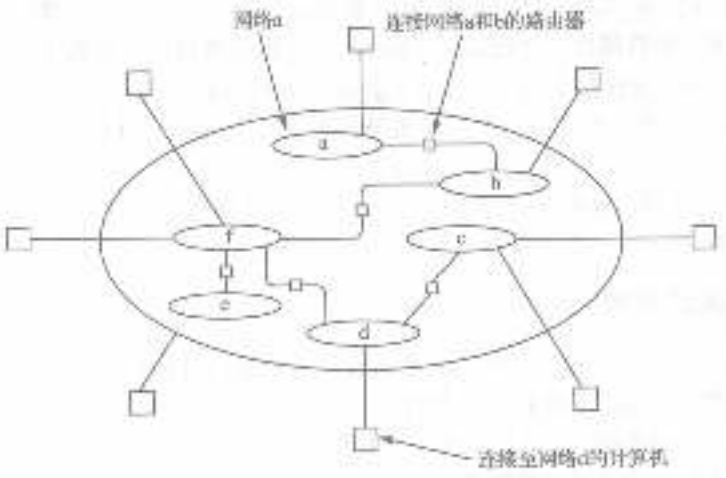


图 15-2　始终驻留在 Internet 内部的物理结构的小例子。每台计算机连接至一个单一网络；路由器将各个网络连接起来

15.8　数据报放在数据包中传输

IP 数据报为所有的 Internet 数据包定义了一个标准的格式。选择一个标准的数据包格式看起来是一个很好的主意，因为这意味着 Internet 数据包不会受到底层网络硬件使用的数据包格式的限制，更为重要的是，这也意味着 Internet 数据包格式不依赖于某种特殊的网络技术。

遗憾的是，定义一个标准的 Internet 数据包格式同样也有缺点。每一种网络技术都定义了它自己的数据包格式，而某一特定的计算机网络只能接收和传送符合自己的网络技术格式的数据包；网络硬件并不能理解 IP 数据报格式。

Internet 如何在不能识别 IP 数据报格式的网络上发送 IP 数据报呢？想像数据报传送的最简单的办法就是考虑隔夜航运服务是如何处理信件的。假设某人写了一封信，把它装入一个信封中，并且在信封外写明收信人的名字，这封信就像一个 IP 数据报。假设寄信人请隔夜航运服务来传送这封信。隔夜服务要求将这封信装入一个他们的信封中，而收信人的名字和地址按照他们规定的格式写在信封外面。外面这层信封就类似于网络数据包（network packet）。

内外两个信封都包含收信人的名字。尽管这两个名字通常是相同的，但是他们并不一定要一致。考虑一下，如果寄信人知道收信人秘书的办公室的确切地址，而不知道收信人的办公室的确切地址，会发生什么呢？寄信人可以将隔夜包裹寄给这个秘书来传递。在这个例子中，内部的地址和外部的地址就是不同的。当这个包裹送达外面信封上的地址时，收信人的秘书打开它并且转寄这封信。

数据包传输使用相同的方式。Internet 把数据报放在一个网络数据包里进行发送。对网络来说，整个 IP 数据报都是数据。当这个网络数据包到达下一台计算机，这台计算机"打开"数据包，取出数据报。接收端检查数据报的目的地址来决定如何处理它。特别地，当路由器决定数据报要被发送到另一个网络时，路由器生成一个新的网络数据包，把这个数据报"封装"到数据包中。然后发送这个数据包使它通过另一个网络到达目的地。当包含数据报的数据包到达它的目的地后，计算机的本地软件打开数据包，然后处理数据报。

15.9　每一台计算机分配一个惟一的地址

为了使数据报的路由和传送成为可能，连接到 Internet 上的每一台计算机都必须分配一个惟

一的地址。正如普通网络中使用的地址一样，Internet 中使用的地址还是数字的。一台计算机在通信前必须知道另一台计算机的地址，正如一个人给另一个人打电话前必须知道他的电话号码一样。

15.10　Internet 地址

分配给一台计算机的惟一的数字称为它的 Internet 地址（Internet address），常缩写为 IP 地址（IP address）。每台计算机（包括路由器）在通过 Internet 通信之前必须被分配一个 IP 地址。当一个组织连接到 Internet 上，它会从相应的 Internet 权威机构获得本组织计算机的一组 IP 地址。如果这个组织添加了一台新的计算机，就从这组地址中选一个分配给这台新机器。

15.11　奇怪的 IP 地址语法

在计算机内部，IP 地址存储在四个字节二进制单元里。虽然这种的内部格式并不重要，但是却能帮助人们理解为什么用一种奇怪的语法来表示 IP 地址。当一个应用程序需要向人们显示 IP 地址，或者人们需要在应用程序中输入 IP 地址时，二进制地址被表示成易于人们理解的形式：用四个被小数点分开的十进制数字来表示。例如，某台特定计算机的 IP 地址为：

128.10.2.1

另一台计算机的 IP 地址为：

192.5.48.3

幸运的是，用户几乎不需要输入或者看见 IP 地址，大多数的应用程序允许人们输入计算机的文字名字来指定一台计算机[⊖]。

15.12　IP 地址不是任意的

在另外一方面 IP 地址也与电话号码类似，这个惟一数字的分配不是任意的。类似于电话系统中的区号，同一个网络上的所有计算机被分配的 IP 地址拥有相同的前缀。地址的合理分配可以更加有效地在 Internet 上路由 IP 数据报。

15.13　示例网络上的传输过程

下面这个例子阐明了 IP 协议软件是如何工作的。考虑图 15-3 所示的网络。

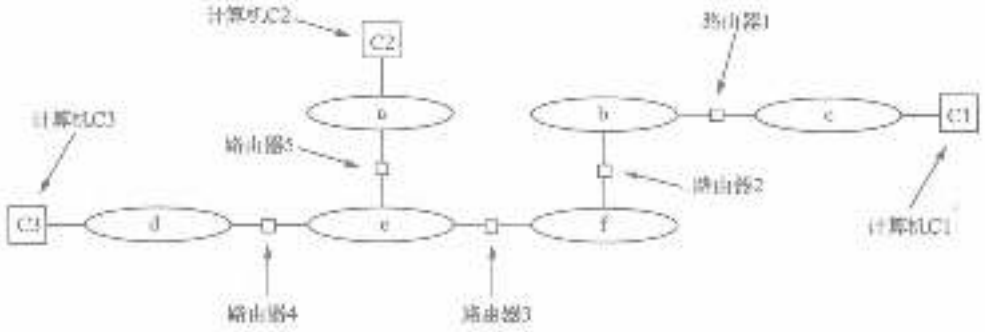


图 15-3　一个示例网络，该拥有六个网络和三台计算机接入

⊖ 第 16 章将介绍计算机文字名字的格式，并解释名字是如何转换为等价的 IP 地址的。

现假设计算机C1需要与计算机C3进行通信。最初，C1上的IP协议软件必须生成一个IP数据报，每个数据报都有两个指明发送端IP地址的域和一个指明目的IP地址的域。C1生成的数据报，把C3的IP地址作为目的地址，C1的IP地址作为发送端（源）地址。

由于这两台计算机并不连接在同一个网络上，所以在它们之间传输的所有数据报必须通过路由器。C1发送的数据报通过网络c到达路由器1。路由器1查找这个数据报的目的地址C3，并决定将它发送到哪里。因为目的地在网络b的另一边，所以路由器1转发这个数据报，通过网络b到达路由器2。路由器2同样查找目的地址并转发它，通过网络f到达路由器3。路由器3必须在路由器4和路由器5之间做一个选择，它选择通过网络e将此数据报发送到路由器4，因为路由器4可以到达目的地[6]。路由器4发现它可以通过网络d把这个数据报传送到最终的目的地C3。如果C3要发送一个数据报返回C1，新的数据报将沿着这条路径的相反方向进行传送。

当然，计算机网络和路由器传送数据报的速度非常快。如果这个例子中的网络使用局域网技术，一个数据报在Internet上传送一个来回只需要千分之几秒的时间。人们感觉这个来回是瞬时的。即使对于远程网络来说，延迟也是非常短的，以至于人们根本察觉不到。所以，Internet能如此有效地传送数据报，使得它运行起来就像一个巨大的单一网络一样。

15.14　你的网络连接有多快

有一个有趣的程序，使用万维网通信来测试连接的速度（带宽）。要运行这个速度测试程序，请根据下面网页的指导来做[7]：

http://www.auditmypc.com/internet-speed-test.asp

这个测试通过双向数据传输来运行：从网站到你的计算机，然后再从你的计算机到网站。在数据传送过程中，软件记录数据传输的速度。尽管此网站测量的是数据在整个Internet上的传输速度，但是数据传输的速度也会根据接入技术（例如，拨号、DSL、电缆调制解调器）的不同而有所不同。这是因为Internet内部通常为更高速的连接。即：

> 接入是Internet中速度的部分——数据传输的速度通常被接入技术运行的速度限制。

这个测试测量两个方向的速度：从网站到你的计算机（称为下行，download），从你的计算机到网站（称为上行，upload）。下行速度和上行速度的不同起因于诸如DSL之类的技术是非对称（asymmetric）的，即，为了更快地显示网页，相对于面向Internet的数据传输，这类技术为面向用户计算机的数据传输提供更多的信道。

15.15　小结

网际协议规定了计算机在Internet上进行通信必须遵守的基本规则。IP协议定义了Internet数据包的格式，即IP数据报的格式。IP协议也定义了地址分配策略，使每台计算机得到在全部通信中使用的惟一的数字。更重要的是，IP协议软件使一组互联的网络和路由器运行起来像一个巨大的单一网络一样。

Internet上的每一台计算机必须安装IP协议软件才能生成和发送IP数据报。每个路由器也安装了IP协议软件，用来将数据报转发到目的地。当数据报到达路由器时，相应的IP协议软件选择一条可以通向数据报目的地的路径。

---

[6] 如果路由器必须在通向目的地的两条路径之间做出选择，那么它会选择最短的一条路径。

[7] 由于这个网站十分受欢迎，所以你可能需要等待一段时间才能进行网络连接速度的测试。

第 16 章　TCP：使通信可靠的软件

16.1　引言

第 15 章讨论了网际协议，并且描述了计算机和路由器上的 IP 协议软件如何使 Internet 上的任意一台机器向另一台机器发送 IP 数据报成为可能。本章继续讨论基本的 Internet 通信软件，研究第二大通信协议：TCP 协议。

16.2　包交换系统会发生过载

第 12 章介绍了多数现代计算机网络使用的基本技术：包交换技术。回忆一下，包交换技术允许多台计算机同时进行通信而没有延迟，因为它要求计算机把数据划分成小的数据包。Internet 中使用的包交换系统，需要另外的通信软件来确保数据传输。为了理解其原因，考虑图 16-1 所示的小型互连网。

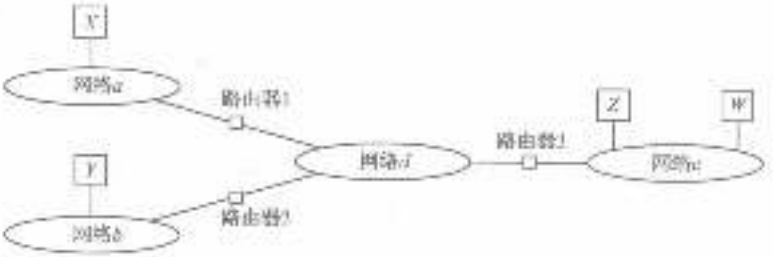


图 16-1　由路由器连接四个网络而形成的互连网示例。如果每个网络的能力相同，那么网络 a 和 b 的数据报是不能全速流向网络 d 的

假设计算机 X 向计算机 W 发送数据包的同时，计算机 Y 向计算机 Z 发送数据包。网络 d 正好位于 X 到 W 和 Y 到 Z 的通路上。另外，这四个网络的能力相同。假设每个网络每秒钟可以处理 5000 个数据包。还假设计算机 X 和 Y 每秒钟生成 5000 个数据报，则路由器 1 和路由器 2 每秒钟各接收到 5000 个数据报。这两个路由器需要将所有它们接收到的数据报通过网络 d 发送给路由器 3。不幸的是，网络 d 每秒钟也只能处理 5000 个数据包。

为了理解这点，想像一下每个网络是一条公路，每个路由器是连接两条公路的交叉口，而所有的公路都有相同的速度限制。图 16-2 说明了汽车在公路上行驶的情形，类似于数据包在互连网上的传输。想像一下，公路 a 和 b 上开满了以限速行驶着的汽车，如果公路 a 和 b 上的所有汽车都试图汇集到公路 d 上，那么一定会发生交通堵塞。

在公路上，当交通堵塞发生时，汽车可以停下来。但在互连网示例中，数据报却不能停下来。每秒钟，从一个网络到达 5000 个数据报，从另一个网络也到达 5000 个数据报，但是只有 5000 个数据报能被发送到它们的目的地。剩下的每秒 5000 个数据报不能到达网络 d，那它们去哪里呢？路由器丢弃它们！当然，每个路由器都有内存，当发生临时阻塞时可以在内存里存储一些数据报。但是，路由器的内存只能够存储几千个数据报。如果数据报仍然以比离开速度更快的

速度到来，在阻塞清除之前，路由器只能丢弃它们。

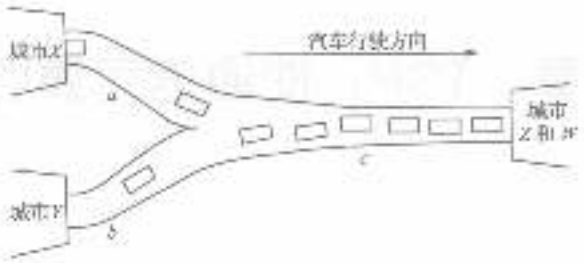


图 16-2　来自两条公路（a 和 b）的汽车汇集到另一条公路（c）上的情形，类似于来自两个网络的数据包汇集到第三个网络上。此图只显示从左向右行驶的汽车，即从城市 X 和 Y 驶向城市 Z 和 W

16.3　TCP 协议帮助 IP 协议确保数据传输

因为 Internet 使用的包交换硬件会发生数据报阻塞，设计者意识到需要另外的通信软件。为了解决这个问题，他们发明了传输控制协议（Transmission Control Protocol，TCP）。TCP 协议使 Internet 更加可靠。所有连接到 Internet 上的计算机运行 IP 协议软件；它们中的大部分计算机也运行 TCP 协议软件。事实上，TCP 协议和 IP 协议都很重要，而且协同工作得很好，所以 Internet 所使用的一整套通信协议一般被称为 TCP/IP 协议族。

TCP 协议解决了在包交换系统中会发生的一些问题。如果路由器发生了数据报过载，它必须丢弃一些数据报。因此，数据报在 Internet 上的传输过程中可能会丢失。TCP 协议自动检查丢失的数据报并处理这个问题。Internet 的结构很复杂，数据报可以在多条路径上进行传输。当路由器或网络系统内的硬件发生故障，其他路由器会在新的路径上传送数据报。这就类似于当高速公路出现交通阻塞时，汽车绕行其他道路的情形，因为路由的路径不同，一些数据报会以与发送时不同的顺序到达目的地。TCP 协议自动检查到达的数据报，并把它们按顺序放好。最后，网络硬件故障会导致数据报的重复传输。因此，同一个数据报的多个副本可能都会到达目的地。TCP 协议自动检查重复的数据报，并只接受第一个到达的数据副本。总之：

> 尽管 IP 协议软件允许计算机发送和接收数据报，但是 IP 协议并不能解决传输过程中可能发生的所有问题。使用 Internet 的计算机还需要 TCP 协议软件提供可靠的，无差错的通信。

16.4　TCP 协议提供计算机程序间的连接

从概念上讲，TCP 协议允许计算机程序间进行交互类似于人们通过电话进行交互。一台计算机上的一个程序指定一个远程程序，并初始化连接（就像拨打电话号码一样）。被呼叫的程序必须接受到达的呼叫（就像接听电话一样）。一旦建立好连接，两个程序间就可以双向地传送数据了（就像进行电话交谈一样）。最后，当两个程序运行结束时，就终止会话（就像挂掉电话一样）。当然，因为计算机以比人们快得多的速度运行，所以两个程序可以在千分之几秒之内建立一个连接，交换少量数据，并终止连接。总之：

> TCP 协议软件使两台计算机可以像人们打电话一样在 Internet 上进行通信。当两个程序建立起一个连接，它们可以交换任意数量的数据，随后终止通信。

16.5 恢复丢失的数据报

检测和删除数据报的副本是一项相对容易的任务。因为 TCP 协议在每个数据报中包含一个数据的标识域，接收端可以用到达的数据报的标识与已经到达的数据标识进行比较。如果是一个数据副本到达，接收端忽略它即可。

恢复丢失的数据报就比较困难了。为了理解其原因，考虑一下在图 16.1 所示的互连网的例子。数据报可能是在互连网中间的某个路由器上丢失的，而最初的源计算机和最终的目的计算机都察觉不到这个问题。TCP 协议使用定时器和应答来解决这个问题。当数据到达它的最终目的地，接收器上的 TCP 协议软件就向源计算机返回一个应答（acknowledgement）。一个应答是一条指明那部分数据已经到达的短消息。

发送端利用应答机制确保所有的数据都到达了。当 TCP 协议软件发送数据时，它利用计算机的内部时钟启动一个定时器。这个定时器就像闹钟一样，当定时器终止，它就通知 TCP 协议。如果应答在定时器终止之前到达，TCP 协议就取消这个定时器。如果定时器在应答到达之前终止，TCP 协议就认为相应的数据报丢失了并重新发送另一个副本。

16.6 TCP 协议的自动重传

许多计算机通信协议都使用下面这个方案：启动一个定时器，如果在定时器终止之前应答没有到达的话，就重传数据。TCP 协议所使用的方案与其他协议有所不同，因为它能够适应 Internet 上各处的要求。如果目的计算机距离源计算机很近（例如，在同一座建筑物内），TCP 协议只需要在重传数据报之前等待较短的时间；如果目的计算机距离源计算机较远（例如，在另一个国家），则 TCP 协议在重传数据报之前要等待较长的时间。另外，超时机制是完全自动的，TCP 协议测量当前 Internet 上的延迟并自动调整超时时间。如果许多计算机同时发送数据报使得 Internet 的速度减慢，TCP 协议就增加重传数据报之前的等待时间。如果情形改变了，Internet 上的数据报传输速度加快了，则 TCP 协议自动减少超时时间。实践表明，在 Internet 上，通信协议必须能够自动地调整超时时间来达到更有效的数据传输。

TCP 协议能够自动调整超时时间的能力，对 Internet 的成功做出了很大贡献。事实上，如果离开了 TCP 协议软件适应条件变化的能力，许多 Internet 应用程序是无法运行的。另外，连续的测量和实践表明，TCP 协议软件能够非常好地适应 Internet 的改变。尽管许多科学家曾经尝试做出改进，但是到目前为止还没有人能够创造出一个工作得更好的协议。

16.7 TCP 协议和 IP 协议协同工作

TCP 协议和 IP 协议能够协同工作得很好并不是一个巧合。尽管这两个协议可以单独使用，但它们是作为一个统一系统的两个组成部分而同时设计的；两者能够互相配合，取长补短。因此，TCP 协议能够处理 IP 协议所不能处理的问题，却并不重复 IP 协议的工作。要点是：

> 连接到 Internet 上的计算机需要 TCP 协议和 IP 协议。IP 协议提供一种将数据从源计算机传输到目的计算机的方法，但是并不处理诸如数据报丢失和乱序传输等问题。TCP 协议处理 IP 协议所不能处理的问题。它们二者一起提供 Internet 上的可靠数据传输。

通常，销售商将 TCP 协议、IP 协议和一些其他相关的通信协议作为一个整体出售。整个软件包称为 TCP/IP 协议软件（TCP/IP software）。

16.8　小结

尽管 IP 协议软件提供基本的 Internet 通信，但是它并不能解决所有可能出现的问题。与其他任何包交换系统一样，如果许多计算机同时发送数据，Internet 也会发生过载。当计算机发送的数据报数目超过 Internet 的处理能力时，路由器必须丢弃一些到达的数据报。

IP 协议软件并不检测丢失的数据报。为了处理这样的通信错误，计算机必须运行 TCP 协议软件。TCP 协议删除重复数据，保证数据按照发送顺序进行重组，并且在数据报丢失时进行重传。

数据丢失的问题很难处理，因为丢失可能在 Internet 上传输的中途发生，甚至接近源计算机和目的计算机的网络和路由器都察觉不到。TCP 协议利用应答和定时器来处理丢失问题。如果在定时器终止之前应答还没有到来，发送方就重传数据。TCP 协议的超时方案能够在整个 Internet 上工作，因为它能够根据目的地距离源的远近来自动调整超时时间。

第 17 章　客户机＋服务器＝分布式计算

17.1　引言

第 16 章介绍了 TCP/IP 通信协议，它们协同工作来提供 Internet 上的可靠数据传输。本章介绍应用程序如何利用 TCP/IP 协议软件在 Internet 上提供服务。本章将说明，尽管 Internet 应用程序多种多样，但是它们都遵循一个组织模型。后面的章节将讨论一些服务实例，并说明这个模型是如何在实际中应用的。

17.2　大型计算机利用网络进行输入输出

早期的计算机体积庞大而且昂贵，因此，大多数公司都只能负担得起一台计算机。当网络技术刚刚出现时，各公司利用这个技术将远程输入/输出设备连接到计算机上。例如，一些远程站点包含一个打印机和几个终端，每个终端又包含一个键盘和一台显示器。尽管远程设备通过网络连接到了中央计算机上，中央计算机却能够完全控制这个设备。这种由一个大型的中央计算机控制远程站点小型输入/输出设备的模式被称为主从式（master-slave）网络。

17.3　小型计算机利用网络进行交互

随着新技术的出现，计算机变得越来越便宜了。微处理器出现后，个人计算机和科研工作站也相继出现。个人计算机虽然体积很小，却不仅仅包含键盘和显示器，它还拥有处理信息的能力。

个人计算机拥有发送和接收数据包所必需的处理能力，所以它能够与网络上的其他计算机直接通信。另外，个人计算机不依赖于大型中央计算机的控制，所以它能够独立工作。为了强调计算机之间相互通信的对称关系，科学家引入术语分布式计算（distributed computing），或称为对等通信（peer-to-peer communication）[1]。

对等通信指允许计算机之间任意通信的网络技术。Internet 支持对等通信是因为它并不区分互相连接的计算机。一台个人计算机可以像它连接一台大型机一样容易地连接另一台个人计算机，一台大型机可以连接另一台大型机、一台中型机，或者一台小型个人计算机。

分布式计算指任何包含两台或多台通过网络进行通信的计算机的计算。这里的计算并不需要包含算术或数字。例如，当两台计算机互相交换电子邮件时，它们就是在进行一种形式的分布式计算，因为这时需要多台计算机协同发送和转发消息。

总之：

> 因为现代计算机网络允许大型计算机或小型计算机发起交互，并且与其他计算机任意地进行交互，所以我们使用术语对等通信。分布式计算指共享访问对等网络的计算机之间的所有交互。

---

[1] 在第 20 章中将会介绍，术语对等（peer-to-peer）被应用于音乐和文件共享应用程序中，这些应用程序在计算机之间进行直接通信。

17.4　Internet 上的分布式计算

Internet 提供种类多样的涉及各种分布式计算的服务。例如，用户可以向朋友发送消息，接收文件、浏览信息、查询数据库、打印文档、发送传真，或者收听音乐。

服务的多样性意味着与 Internet 交互方式的多样性。在一些情况下，用户与其他用户进行交互。另一些情况下，用户与提供服务的远程计算机程序进行交互。还有一些情况下，两个计算机系统进行通信，而不需要人的干涉。交互式服务允许用户保持几个小时或者几天的连接。其他服务则仅需几毫秒来提供所需信息，并且几乎立即终止通信。一些服务允许用户获取信息，而另一些服务允许用户存储和更新信息。一些服务只涉及两个计算机系统，一个发送请求而另一个提供应答；另一些服务则涉及多台计算机。

17.5　通过一个范例解释所有分布式计算

尽管 Internet 服务多种多样，而且它们的用途区别很大，但是实现服务的软件总是使用一种方案。这个方案称为客户机/服务器计算（client server computing）。客户机/服务器计算并不局限于 Internet，它是所有分布式计算的基础。

客户机/服务器计算的思想非常简单：Internet 上的一些计算机提供服务，而另一些计算机访问服务。例如，一些服务器管理存放信息的文件，客户机程序可以连接这些服务器来请求某个文件的副本。另一些服务器管理多用户游戏，想加入某一游戏的用户必须让客户机程序连接到服务器才能参加该游戏。

要理解为什么一个范例可以涵盖服务的多样性，必须了解以下三点：

- 程序通信。

  Internet 的用户经常说他们的计算机与别的计算机进行了通信。尽管这经常出现在非正式的交谈中，但它隐藏了一个技术细节。计算机之间并不能进行通信，只有程序之间才可以通信。运行在一台计算机上的程序利用协议软件去连接另一台计算机上的程序，并交换信息。在 Internet 上，这两个程序必须使用 TCP/IP 协议软件。尽管计算机和运行在计算机上的程序之间的差别是微不足道的，但这些差别却是很重要的，因为它解释了为什么一台计算机可以与多台计算机进行会话。

- TCP/IP 协议并不创建或运行应用程序。

  虽然 Internet 可以将数据从一点传送到另一点，但是它并不能自动启动接收方机器的程序。在某种意义上，Internet 像电话系统一样，它允许一个程序呼叫另一个程序，但是在通信进行之前被呼叫程序必须做出应答。因此，两个程序只有在其中一个程序启动运行并且在其他程序连接它之前同意应答呼叫的情况下才能运行通信。

- 计算机可以运行多个程序。

  即使最慢、最小的计算机也能在同一时间运行多个程序。多数的计算机只有一个处理器，所以一台计算机能够运行多个程序看起来就很奇怪。但是，操作系统在多个程序之间快速地切换处理器使得它们可以同时运行。操作系统允许处理器在一个程序上运行一小段时间，然后切换到另一个程序上运行一小段时间，依此类推。因为计算机的处理器能够每秒执行几百万个操作，所以在多个程序之间切换处理器就给了人们所有的程序在同时运行的感觉。例如，用户可以在显示器的三个独立窗口上运行三个任务，各个任务看上去都在同时运行。

了解了程序通信和计算机可以同时运行多个程序之后，就可以理解为什么一台计算机可以

同时为多个用户提供服务。许多Internet服务使用一个程序的多个副本来允许多个用户同时访问这项服务。例如，一台计算机可以同时接收和存储从其他多台计算机发送来的电子邮件。为了做到这一点，它创建接收电子邮件程序的多个副本，每台发送邮件的计算机与该程序的一个副本进行通信。由于处理器可以在多个副本间快速地切换，因此所有的通信看上去是同时进行的。

总之：

> Internet上的通信总是发生在一对程序之间：一个程序初始化会话，另一个程序必须等待并接收它。因为一台计算机可以同时运行多个程序，所以它看上去可以同时进行多个会话。

17.6 通信程序分为客户机和服务器

进行通信的计算机程序可以被分成两类：所有提供服务的程序属于服务器（server）类；所有访问服务的程序属于客户机（client）类。

通常，Internet的用户使用客户机软件。例如，一个利用Internet访问一项服务的应用程序就是一个客户机。这个客户机利用Internet与一个服务器进行通信。在一些服务中，客户机使用一个请求来与服务器进行交互。客户机生成一个请求，把它发送给服务器，并等待应答。在另外一些服务中，客户机需要长时间地与服务器进行交互。客户机建立与服务器的通信，将键盘和鼠标的输入传送给服务器，并不断地显示来自服务器的数据。

17.7 服务器必须一直运行

与客户机软件不同，服务器程序必须随时准备接收请求。客户机可以在任何时候向服务器发出请求；服务器并不能提前知道请求发出的时间。通常，服务器程序运行在允许多个服务器程序同时执行的大型计算机上。系统启动以后，每个服务器程序的一个或者多个副本开始运行。只要计算机在运行，服务器就一直在运行。

如果计算机掉电或者操作系统崩溃，运行在该计算机上的所有的服务器将会消失。如果提供服务的计算机崩溃，使用该计算机上的服务器的活动的客户机将会接受到一条错误信息。当该计算机停止运行后，任何试图与服务器建立通信的客户机也将收到一条错误消息。

17.8 小结

与过去的网络将终端连接到大型的中央计算机上不同，Internet提供对等联网技术。它允许任意的计算机与其他计算机进行通信。

Internet提供多种使用不同交互方式的服务。尽管各个服务之间有着明显的差别，但是Internet上的所有软件使用了相同的基本结构。这个结构被称为客户机/服务器计算。

在客户机/服务器环境中，每个程序必须成为客户机或服务器之一。服务器程序提供服务。通常，提供服务的计算机在启动时就开始自动运行服务器软件。服务器一直运行，随时准备接收到来的请求。当用户访问服务时通常运行客户机软件。客户机程序连接服务器程序，发送请求，并显示服务器程序的应答。

第 18 章　计算机的命名

18.1　引言

第 17 章介绍了提供和使用服务的计算机程序之间的通信。本章将介绍 Internet 的一个重要服务，使人们能用文字名字代替数字地址来表示计算机，还将说明命名方案，并介绍软件如何利用客户机/服务器方式的交互来将计算机的文字名字转换成数字地址。

18.2　比起数字来人们更喜欢用名字

回忆一下，Internet 分配给每台计算机一个称为 IP 地址的数值，并且每个通过 Internet 传输的数据包都包含目的计算机的 IP 地址。像电话号码一样，IP 地址也包含多位数字，因此很难记忆并保证准确输入。例如，如果用十进制表示，IP 地址包含十位数字。

比起 IP 地址来，人们更喜欢使用计算机的文字名字。Internet 接受这种文字名字。首先，Internet 允许用户为自己的计算机命名。其次，软件允许用户输入计算机的文字名字来代替 IP 地址。最后，Internet 包含一项服务自动地将计算机的文字名字转换成数字地址。我们在介绍了计算机的名字是如何分配的之后，再讨论名字转换的问题。

18.3　为计算机命名既困难又有趣

为计算机分配的名字能够影响人们对它的印象；名字代表个性。例如，一些人为他们的计算机选择的名字表达了他们在工作中的成就感，而另一些人选择的名字则表达了他们的挫败感。当人们听到一台计算机的名字时，往往会推断出关于这台计算机或它的拥有者的一些信息。

Internet 上计算机的名字多种多样。例如，人们选择地理位置、电影角色、演员、颜色、海洋、公司、神话人物、名人的名字为自己的计算机命名。如“Mercury”（水星）、“Saturn”（土星）等行星的名字一直很流行。1996 年，除了如“mailhost”（邮件主机）之类普通的名字，Internet 上最流行的计算机名字是“venus”（维纳斯）；到了 2000 年，“mars”（战神）超过了“venus”；但是到了 2005 年，“zeus”（宙斯）超过了“mars”。

当然，使用普通的词作为计算机的名字看起来过于随意了。有些人为他们的计算机选择了能够识别其型号的名字，并且仅简单地加一个后缀来区别多台机器。例如，pc1、pc2……系列，host1、host2……系列等的使用在 Internet 上排在前列。

一组合作者可以为他们的计算机选取一种命名方案。例如，一组人使用树木的名字，他们的计算机被命名为“birch”（桦树）、“elm”（榆树）、“oak”（橡树）等。通常，命名方案与组员的职业有关。一群化学家的计算机可以用元素来命名（例如，“hydrogen”（氢）、“helium”（氦）和“oxygen”（氧）），一群物理学家的计算机可以用粒子来命名（例如，“proton”（质子）、“quark”（夸克）、“neutrino”（微中子）和“positron”（正电子））。

有些人觉得为他们的计算机命名十分有趣。在一所大学里，计算机被命名为“up”（上）和

"down"（下），使得"今天上变成下了"[1]这句话成真了。漫画人物的名字也非常流行。你可能会觉得奇怪，"snoopy"（史努比）在 Internet 上的计算机名字中排在前几百名；同时，"calvin"排在前 50 名，而"Calvin and Hobbes"（卡尔文与霍布斯）是一部非常流行的漫画。有时计算机和它的拥有者使用相同的名字，可以用它来和这台计算机或这个人开玩笑。例如，如果计算机和它的拥有者都叫做"John"（约翰），别人就能说"约翰今天关机了"或者"我觉得很难与约翰一起工作"。

除了为了乐趣选择的名字之外，一些计算机的名字能够反映它们的主要角色。运行着某个组织电子邮件服务的计算机常常被命名为"mail"（邮件）或"smtp"（用来发送邮件的协议）。类似地，运行着某个组织的万维网服务的计算机常常被命名为"www"或"Web"。图 18-1 表明，在 2005 年，许多最常使用的计算机的名字都是服务名称[2]。使用服务名称能够使人们更方便地运行和维护组织的计算机。

排名	名字	排名	名字	排名	名字
1	www	18	www2	35	web
2	mail	19	webmail	36	fw
3	dsl	20	mailgate	37	dns1
4	cpe	21	network	38	host1
5	ns2	22	dns	39	ns3
6	ns	23	mail1	40	e0
7	ns1	24	pc3	41	dns2
8	gw	25	pc4	42	host2
9	ftp	26	secure	43	admin
10	router	27	pc5	44	vpn
11	smtp	28	mailhost	45	mx
12	pc1	29	proxy	46	bcast
13	pc2	30	test	47	server1
14	mail2	31	demon-gw	48	www1
15	gateway	32	pc6	49	szerver4
16	server	33	gate	50	szerver3
17	broadcast	34	net		

图 18-1 2005 年 Internet 上 50 个最常用的计算机名字

18.4 计算机的名字必须惟一

尽管大多数人都喜欢使用简短的名字来命名他们的计算机，但是在 Internet 上必须要使用长名字来避免给多台计算机分配相同的名字。两台计算机使用相同的名字将会产生严重的问题，因为通信软件将不能区分它们。要点是：

> Internet 上的每台计算机都必须有一个惟一的名字，否则就不能将这台计算机和其他计算机区分开来。

18.5 计算机名字的后缀

为了使计算机的名字是惟一的，Internet 的命名机制使用了一种常见的思想：在每一个名字后面添加额外的字符串。我们把这些额外的字符串看成是为名字添加的后缀。计算机的全名包

[1] 原文是"down is up today"，——编者注

[2] 以后会有章节介绍各种 Internet 服务，包括文件传输服务（File Transfer Protocol，FTP）和万维网（World Wide Web，WWW）。

括它的本机名加上一个小数点，然后再附加上组织后缀。

人类名字的附加部分表示一个人的家族或者出生地，Internet 名字的附加部分则表示拥有该计算机的组织。然而，有时多个组织拥有相似的或者相关的名字。为了使这么多组织可以无冲突地使用相似的名字，Internet 命名方案进一步规定了名字后面要添加组织的类型。例如，Internet 权威机构将 Purdue 大学分为教育机构类，所以 Purdue 大学所有的计算机都要添加后缀：

purdue.edu

如果一个名叫"Purdue Gumball"的公司要在它的计算机的名字里使用"Purdue"，Internet 命名权威机构将会分配给它一个表明它是商业企业的后缀。例如，它可能被分到：

purdue.com

后缀"purdue.com"清楚地将该公司和 Purdue 大学区分开来。

如果 Purdue 大学和"Purdue Gumball"公司各自命名一台计算机为"groucho"，则后缀能够保证这两台计算机都有惟一的名字[⊖]：

groucho.purdue.edu

和

groucho.purdue.com

要点是：

> 因为添加到一台计算机名字后面的后缀表示拥有这台计算机的组织和该组织的类型，所以能够保证被不同组织拥有的两台计算机的全名是不同的。

18.6　名字由多部分组成

尽管上面的例子说明计算机的名字通常包含三个部分：本机名、组织名和组织类型，但还不限于此。Internet 命名方案允许计算机的名字包含多个部分。Internet 分配给某个组织后缀以后，该组织还能够为它的计算机的名字添加额外的部分。通常，每个组织的所有计算机的名字包含本机名、部门名称以及组织后缀。在名字里添加额外的一层可以解决一个重要问题：

> 因为 Internet 允许各组织在计算机名字后面添加额外的部分，所以组织里的每个小组都可以自由地为该组的计算机选择本机名。可以保证在不同小组的任意两台计算机的全名都是不同的。

与许多其他大学一样，Purdue 大学的计算机科学系使用两个字母的缩写"cs"来表示该系。因此，"cs.purdue.edu"就作为该系所有计算机的名字的后缀。例如，Purdue 大学计算机科学系的一台名叫"groucho"的计算机的全名是：

groucho.cs.purdue.edu

类似地，该系的一台名叫"www"的计算机的全名是：

www.cs.purdue.edu

Purdue 大学电气与计算机工程系使用缩写"ece"，因此该系的一台名叫"www"的计算机的全名是：

www.ece.purdue.edu

18.7　美国之外的计算机域名

尽管美国的大部分计算机域名都是以".edu"或者".com"结尾的，但是其他的国家也可

⊖ 在 Internet 术语中，一台计算机被称为一台主机（host），计算机的全名更称为主机名（hostname）。

以选择不同的方案。其他国家通常在域名后添加两个字母的国家代码。例如在中国，雅虎网站被命名为

yahoo.com.cn

因为“cn”是国际认可的表示中国的两个字母的国家代码。类似地，德国的计算机域名以“de”结尾，加拿大的计算机域名以“ca”结尾，英国的计算机域名以“uk”结尾。每个国家都能选择如何进一步划分域名。例如，因为英国选择“ac”来表示学术机构，所以英格兰 York 大学的一台计算机可以被命名为：

minster.cs.york.ac.uk

18.8 将计算机的名字转换成对应的 IP 地址

前文提到 Internet 通信软件在发送和接收数据报时必须使用 IP 地址。尽管人们可以用名字来表示一台计算机，但是在可以向该计算机发送数据包之前必须把它的名字转换成 IP 地址。Internet 有一项自动将计算机的名字转换为 IP 地址的服务。前面所描述的文字名字被称为*域名*（domain name）。将域名转换为 IP 地址的软件被称为*域名系统*（Domain Name System，DNS）。当应用程序碰到一个计算机域名（例如，当用户从键盘输入一台计算机的域名）时，它利用域名系统将该域名转换成 IP 地址，然后在后面的通信中使用该 IP 地址。

域名系统使用第 17 章介绍的客户机/服务器模式。各组织都运行一个包含该组织所有计算机和对应的 IP 地址的域名服务器。当应用程序需要将某台计算机的域名转换成 IP 地址的时候，该应用程序成为 DNS 的客户机，它连接域名服务器，然后将该计算机的域名发送给该服务器。发送计算机的域名给域名服务器相当于询问：

这个域名对应的 IP 地址是什么？

该服务器查线答案，然后将正确的 IP 地址返回给客户。

18.9 域名系统的工作方式类似于电话系统的目录帮助服务

理解域名系统工作原理的最简单的办法就是思考一下电话系统的目录帮助服务。要拨打一个电话，拨打电话的人必须知道对方的电话号码。如果拨打者知道对方的全名、地址和城市，那他就可以通过目录帮助服务查询到对方的电话号码。查询时，拨打者呼叫对方所在城市的目录帮助服务，并说明对方的名字。目录帮助服务的接线员则返回所查询的人的电话号码作为应答。

当然，一个目录帮助服务的接线员不可能知道世界上所有的电话号码。例如，如果一个人需要为一个国家某人的电话号码，而本地的目录帮助服务不能提供答案。但是，本地的目录帮助服务却能够提供外国的目录帮助服务的电话号码。

域名系统的工作方式与目录帮助服务类似。一台指定的域名服务器不可能存储 Internet 上所有计算机的域名和对应的地址。实际上，每台域名服务器只存储本地计算机（例如，某个公司或企业的计算机）的域名。当法国某台计算机的某个程序需要知道加利福尼亚的一台计算机的 IP 地址时，该应用程序将它的请求发送给法国的域名服务器。法国的域名服务器不知道答案，但是却知道如何与加利福尼亚的适当的域名服务器联系。

18.10 计算机名字的查询是自动的

考虑以下这个例子。假设法国的一个用户想要与加利福尼亚 XYZ 公司的一台名叫“calvin”的计算机通信。这个用户必须指明该计算机的域名，例如，该计算机的全名可能是：

calvin.xyz.com

在法国的计算机上的某个应用程序能与计算机"calvin"通信之前，它必须得到"calvin"的IP地址。为了得到这个地址，该应用程序使用Internet的域名系统。

法国的应用程序需要知道XYZ公司的域名服务器地址吗？不需要，域名系统是完全自动的，Internet上的计算机只需要知道一台域名服务器的地址，通常是本地域名服务器。域名服务器自动地处理查询操作。图18-2说明了这一通信过程。

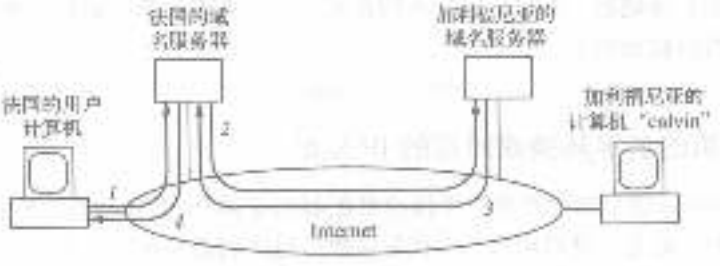


图18-2 为了与远程计算机进行通信，应用程序向本地域名服务器查询远程计算机的IP地址。如果本地域名服务器不知道答案，它自动地联系远程域名服务器。箭头上的数字说明了四个步骤的顺序

如图18-2所示，查询远程计算机IP地址需要四个步骤。第一步，位于法国的计算机上的应用程序向本地域名服务器发送一条消息，其中包含要查询的域名calvin.xyz.com。该应用程序向域名服务器询问：

XYZ公司的名叫"calvin"的计算机的IP地址是什么？

第二步，法国的本地域名服务器与XYZ公司的域名服务器联系。第三步，XYZ公司的域名服务器发送一个应答。第四步，法国的本地域名服务器将答案返回给最初的应用程序。注意，域名服务执行这四个步骤时并没有联系计算机"calvin"，就像目录帮助服务可以提供某人的电话号码而不用拨打电话给他一样。

尽管域名系统需要通过Internet发送几条消息，但是得到应答却不需要等待很长时间。在大多数情况下，不到一秒钟就能得到应答。速度非常重要，因为在用户输入一台需要与之通信的计算机的域名时，往往需要进行域名查询。当域名系统查找计算机的IP地址时，用户必须等待。只要得到了域名系统的应答，应用程序就能够开始直接向该计算机发送数据包了[⊖]。

总之：

> Internet上的计算机只需要知道一台域名服务器的地址。当一个程序向本地的域名服务器发送某台计算机的域名时，它就成为域名服务器的一个客户机。本地域名服务器将域名转换成IP地址进行应答。尽管本地的域名服务器可能需要联系远程的域名服务器才能获得信息，但是客户机总是从本地域名服务器获得答案。

18.11 IP地址与域名并不相关

在语法构成上，域名与IP地址非常相似。IP地址是由被三个小数点分隔开的四个数字组成的，而域名则由被小数点分开的字符串组成。

表面现象通常具有欺骗性。尽管它们看起来很相似，但是域名和IP地址的各个部分是完全

⊖ 为了进一步优化速度，在计算机查询了某个域名后，会在短时间内保留其结果。因此，如果用户连续访问第三次同样的网站，那么只需要在第一次访问时进行域名查询。

不相关的。IP地址包括四个部分，是因为32位的二进制数字可以被分割成四个8位的部分。域名包括多个部分，是因为一个组织可以在它的后缀里不添加或者添加多个项，使得组织内部不同的小组可以为它们的计算机选择相同的本机名。

由于有些域名也正好包括四个部分，这就容易产生混淆。例如，Internet的域名服务将域名：

arthur.cs.purdue.edu

转换成IP地址：

128.10.2.1

但是，字符串"arthur"与"128"并不相关，域名与IP地址的其他对应部分也是如此。为了弄清楚这个区别，考虑一下一个人的名字和他的电话号码。某人的名字中的字符与他的电话号码中的数字毫无关系（例如，某人的姓和他的电话号码中的第一组数字毫无关系）。

18.12 小结

由于相对于数字来说，人们更喜欢使用名字，所以Internet允许为每台计算机分配一个名字，并且提供了一个能自动地将名字转换成对应的IP地址的系统。

许多Internet上的应用程序允许用户输入远程计算机的域名。应用程序向本地的域名服务器发送一条包含要查询的域名的请求消息，成为域名服务器的一个客户机。域名服务器将该域名转换成IP地址，并且在需要的情况下自动地与其他的域名服务器联系。

计算机的域名包含多个由小数点分割开的字符串。计算机的本机名在最前面，最后面的后缀表示拥有这台计算机的组织。如果该组织还添加了表示各部门信息的名字，那么域名中间还要添加额外的字符串。

尽管计算机的域名和IP地址都是由被小数点分隔开的字符组成的，但是域名中的各项与IP地址中的各项毫不相关，就像一个人名字中的字符与他的电话号码中的数字毫不相关一样。

第 19 章　NAT：共享 Internet 连接

19.1　引言

第 14 章介绍了 DSL 和电缆调制解调器等提供宽带访问 Internet 的技术。本章将介绍对基本访问技术的扩展，一种允许一组计算机“共享”一个物理连接和一个 IP 地址的机制。本章将说明共享的动机以及实现共享的技术。

19.2　大容量与多台计算机

第 14 章说明了 DSL 和电缆调制解调器技术提供高速、专用的 Internet 访问。在各情形中，底层技术提供用户与 ISP 之间的专用连接。宽带连接的一大好处在于其速度：宽带连接可以每秒传送大量数据，其连接容量足够多台计算机使用。

然而，许多 ISP 设计的宽带系统有一个限制：一个用户每次只能使用一台计算机。举个例子，前面提到 Internet 上的每台计算机都需要一个 IP 地址。许多 ISP 网络只分配给每个用户一个 IP 地址，申请更多的 IP 地址的请求将会被拒绝。所以，仅仅将更多的计算机连接到 DSL 或电缆调制解调器上是不能正常工作的：只有第一台发出请求的计算机能够获得一个 IP 地址，其他的计算机会被阻止使用 Internet。

19.3　共享一个 IP 地址是可能的

由于 ISP 的宽带服务限制每个用户只能使用一个 IP 地址，Internet 的研究者就开始考虑安排位置相近的多台计算机“共享”一个 IP 地址是否可能。如果能够发明一种共享机制，每个用户就能同时使用多台计算机（例如，当父母阅读邮件的时候，孩子们可以查询学校信息）。

起初，共享连接的想法似乎很荒谬。毕竟，整个 Internet 的设计就是每台计算机分配惟一的 IP 地址。如果多台计算机使用同一个 IP 地址，数据报如何才能被传送到正确的计算机呢？一些工程师断言，共享连接违背了 Internet 的基本设计原则。

聪明的设计师发明了一种方法来解决这个问题：不要把计算机直接连接到 DSL 或者电缆调制解调器上，而是在它们中间插接另外一个设备。这个额外设备可以连接多台计算机共用一个 IP 地址。

19.4　实现共享连接的设备称为 NAT 盒

允许连接共享的设备被称为网络地址转换（Network Address Translation，NAT）设备；业内使用非正式术语 NAT 盒（NAT box），图 19-1 说明了 NAT 设备是如何将多台计算机和宽带调制解调器连接起来的。

要点是：

> 一种被非正式地称为 NAT 盒的设备允许用户在多台计算机之间共享一个 Internet 连接。共享在宽带连接的情况下才是最可行的，因为宽带连接有足够的容量来支持多台计算机。

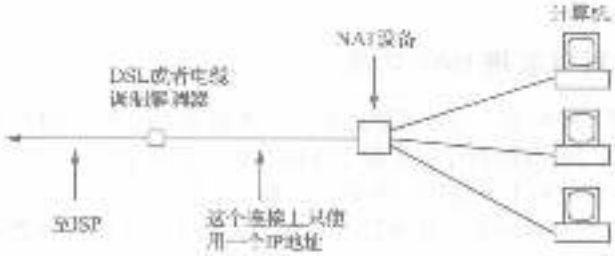


图19-1 允许多台计算机共享一个IP地址的物理连接方式。NAT设备处于宽带调制解调器和一组计算机之间

19.5 NAT盒的行为与小型ISP类似

NAT设备使用户的ISP和连接到NAT盒上的计算机都产生一种幻觉。对ISP来说，NAT盒看起来像一台单独的计算机，当NAT盒启动之后，它就完全像一台计算机启动了一样进行通信：NAT盒通过DSL或者电缆调制解调器向ISP发送数据报，并且请求一个IP地址供未来所有的通信使用。因此，被请求的ISP认为这个用户只有一台计算机。

对一台计算机来说，NAT盒看起来像一个ISP。当某台计算机启动之后，它请求一个IP地址，NAT盒就发送一个IP地址给它作为响应。有趣的是，NAT盒分配给计算机的IP地址在Internet上是无效的。NAT盒使用一种本地地址，它被限制在计算机和NAT设备之间的链路上。一组特殊的IP地址就是为这样的目的而保留的：NAT设备通常使用以下格式的IP地址：

192.168.1.*xxx*

其中*xxx*是为连接到NAT盒上的每台计算机所分配的惟一的数字。例如，第一台请求IP地址的计算机可能被分配IP地址：

192.168.1.100

下一台计算机可能被分配IP地址：

192.168.1.101

等等。

19.6 NAT改变每个数据报中的地址

从本质上讲，NAT盒的行为与路由器类似，它能够在计算机和Internet之间转发数据报。其实，出售NAT设备的厂商通常称其为路由器（router）。NAT可以只在它连接的计算机和Internet之间路由数据报吗？答案是否定的，因为NAT连接的计算机使用的IP地址在Internet上是无效的。为了实现其功能，NAT盒必须改变每个数据报中的IP地址。实际上，在数据报被发送到Internet之前，NAT盒将源地址替换成NAT盒知道的有效的IP地址。当Internet上的某个站点进行回复时，这个回复将会被转发给该NAT盒。

如果所有发送出去的数据报都有一样的回复地址，NAT盒如何才能知道从Internet来的数据报该被哪台计算机接收呢？实际上，NAT盒对每台计算机正在做什么（例如，正在访问某个网站或者正在下载某份文件）保留一份记录。当一个数据报从Internet到来时，NAT盒参考这一记录来决定数据报该由哪台计算机接收。在转发之前，NAT盒将该数据报的目的地址替换成分配给目标计算机的本地地址。

总结如下：

> 尽管网络地址转换设备与路由器类似，而厂商也通常将这种商业产品称为路由器，但是NAT盒与普通路由器是不同的，因为NAT盒对数据报中的IP地

进行转换。

19.7　计算机软件可以实现 NAT 功能

当前存在很多 NAT 的版本，最易理解的形式可参见图 19-1 所示的由小型物理设备组成的连接。但是，编写一个专门的应用程序以实现 NAT 功能也是可行的。在 Linux 操作系统上运行的免费的 NAT 软件就在学生中十分流行。例如，如果一个宿舍只有一个 Internet 连接，那么可以为其中一台计算机另外买一块网卡，并在该计算机上安装 Linux 操作系统和 NAT 软件，就可以连接其他的计算机了。

从逻辑上讲，软件版本的 NAT 与图 19-1 所示的结构并没有区别。但是，物理连接有区别，这里需要小型、廉价的网络集线器（network hub）提供物理连接。一台运行 NAT 软件的计算机有一条连接到集线器的附加的网络连接。如图 19-2 所示。

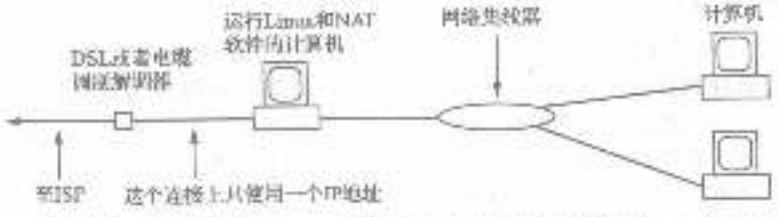


图 19-2　当 NAT 软件在一台计算机的 Linux 操作系统上运行时的连接情形。从逻辑上讲，与图 19-1 的结果是相同的

除了 Linux 操作系统可以运行 NAT 软件，还有适用于 Microsoft Windows 和 Mac 机上的 Apple OS X 操作系统的各种 NAT 软件版本。Microsoft 的版本被称为 Internet 共享连接（Internet Connection Sharing），Mac 的版本则是 airport 的一部分。其中每个版本所提供的功能在本质上与 Linux 版本是一致的。

19.8　NAT 可以用于无线网络

商业 NAT 产品最有趣的应用之一是涉及无线网络技术的使用，通常是 Wi-Fi 的应用。许多厂商在做广告时将 NAT 设备称为无线路由器（wireless router）。在本质上，无线路由器仅仅扩展了 NAT 的概念，即在允许有线连接的同时允许 Wi-Fi 连接。例如，见图 19-1。想像一下，除了图中所示的三台计算机，包含 Wi-Fi 转发器的 NAT 设备可以使更多的计算机进行通信。

第 31 章将要讨论由无线路由器引起的一个有趣的问题。因为它们运行 NAT，所以这些路由器可以将任意一台无线计算机连接到 Internet 上，一台配置不当的无线路由器有可能使附近（例如，相邻的公寓）的某人连接到 Internet，并共享某用户的连接。

19.9　小结

尽管 ISP 通常限制每个用户只能拥有一个 IP 地址，网络地址转换（Network Address Translation，NAT）技术却允许用户在一组计算机之间共享一个连接。共享在宽带连接的条件下最有意义，因为宽带访问有足够的容量来处理来自多台计算机的通信。

对 ISP 来说，NAT 设备就像一台计算机一样，对每台计算机来说，NAT 盒就像一个 ISP。当数据报通过时，NAT 盒在本地 IP 地址和惟一的外部地址之间进行转换，所以 ISP 察觉不到多台计算机的存在，同时这些计算机也察觉不到它们在共享一个连接。

除了它们在有线网络连接中的使用之外，NAT 盒也能提供无线网络连接。正确的配置是必需的，以保证无线服务没有在无意中延伸到邻居。

第 20 章　为何 Internet 运行良好

20.1　引言

前面几章介绍了基本的 Internet 技术，包括 TCP/IP 协议软件。这一章讨论一下 Internet 成功的原因，以及值得借鉴的地方。

20.2　Internet 运行良好

Internet 是一个技术成功的奇迹。基本的 TCP/IP 技术能够适应最初的设计者所没有预料到的发展和变化。当 Internet 上的计算机数量呈指数级增长时，TCP/IP 技术能够适应规模的扩张。同时，Internet 上的通信量也呈指数级增长，TCP/IP 技术同样能够处理增加的数据包。尽管现在计算机的运行速度比 TCP/IP 协议刚刚诞生时快几千倍，但是新式计算机能够通过 Internet 互相通信，也仍然能与老式计算机进行通信。尽管 Internet 中心的速度增加了 150 000 倍，但是基本的 TCP/IP 协议并没有改变；同样的设计仍然能够在更高的速度下正确运行。

为什么 TCP/IP 技术能够如此成功？一个研究项目中产生的技术是如何成为世界最大的计算机网络系统的基础的呢？我们能从 Internet 项目中学到些什么？显然，任何一项单独的技术决策并不能完全决定像 Internet 这样的复杂系统的成功。但是，一个错误的设计选择却会毁掉一个出色的计划。本章以下部分将讨论 TCP/IP 中一些优秀的设计选择，并从 Internet 项目中汲取一些经验。

20.3　IP 协议提供灵活性

IP 协议提供了适应范围广泛的底层网络硬件所必需的灵活性。例如，IP 协议可以利用：

- 广域网技术或局域网技术。
- 以最高速度运行的网络或以最低速度运行的网络。
- 保证不丢失数据包的网络或提供尽力而为服务的网络。
- 使用无线电进行通信的无线网络，使用金属导线传送信号的网络，或使用光纤传送信号的网络。
- 以上各种网络的组合。

总之，IP 协议允许 Internet 利用几乎任何类型的计算机通信技术。

IP 协议成功的秘诀在于允许各种类型网络技术的存在。因为对网络硬件的要求不高，所以 IP 协议几乎支持任何一种将数据从一处传送到另一处的机制。从设计角度来说：

> IP 协议适用于许多硬件类型，是因为它几乎不对底层的网络硬件进行任何设定。

尽管 IP 协议对网络硬件作了最少的设定，但是 IP 协议的所有实现必须遵守完全相同的通信规则。为了保证各种 TCP/IP 实现的相容性，人们将协议的完整的规范记录在文档中，并非正式地称之为标准（standard）。TCP/IP 协议标准包括严格的规范来说明如何在各种类型的网络上传送 IP 数据报。当一种新的网络技术出现后，一个新的 Internet 标准文档将会阐述如何与该技术

一起使用TCP/IP协议，这些规范是Internet文献的一个重要部分，因为它们保证所有的计算机和路由器在网络上发送数据报时都使用完全相同的格式。因此，

> 因为TCP/IP标准文档严格地规定了如何在一种给定类型的网络上传送IP数据报，因此多数厂商提供的计算机和路由器都遵守相应的细节。

20.4 TCP协议提供可靠性

TCP协议和IP协议互为补充保证网络的良好运行。TCP协议处理IP协议所没有处理的通信问题，并且为应用程序提供可靠的通信。

有趣的是，TCP协议必须能够应付IP协议所能使用的各种网络硬件类型之间的差别。例如，通过卫星传送一个数据报需要几十秒，而通过局域网传送一个数据报只需要几毫秒。TCP协议软件必须能够处理这两种情况。同样地，局域网中几乎或者从不发生丢包现象，而广域网中会丢失一定比例的数据包。TCP协议软件必须能够高效地使用这两种技术。

TCP协议还能够处理包交换系统中最困难的问题：性能的快速变化。计算机常常突发地传送信息，计算机在一段时间内保持安静，然后在短时间内发送大量数据，然后再恢复到安静状态。例如，当用户启动一个应用程序以后，该应用程序需要与服务器进行交互（例如，下载一个文件或者获得登录信息）。当用户停下来思考问题、移动鼠标或者通过键盘输入数据时，应用程序就停止与服务器的通信。尽管Internet有足够的能力来处理多台计算机间发送的数据报，但是如果许多计算机同时发送大量数据，Internet就会临时过载并且速度减慢。TCP协议能够监控延时现象，并在拥塞消除之前进行等待。

TCP协议成功的秘诀在于它能够自动适应网络的变化。

> TCP协议能够随时监控Internet条件的变化并进行自动调节，因而使Internet在发生临时拥塞时也能够进行可靠的通信。

20.5 TCP/IP协议软件的开发注重高效性

在任何复杂的计算机系统中，设计者必须在许多可能的设计中做出选择。TCP/IP协议在设计时充分考虑了运行的高效性。例如，TCP/IP协议在发送和接收数据包之前不需要进行大量计算。另外，TCP/IP协议只发送满足通信所需的最少的网络数据包。

这些高效的设计使TCP/IP协议软件能够在低速的小型计算机上运行，也能够在高速的大型计算机上运行。因此，TCP/IP协议即使在没有像大型计算机那么大处理能力和内存的个人计算机上也能很好地工作。

20.6 TCP/IP协议的研究着重于实际结果

从事Internet项目的科学家和工程师们采用实用的方法来进行研究。他们并不进行抽象的讨论，而是建立、测试、衡量了一个实际可运行的通信系统。他们通过实验现象来评价所有新的建议和想法。例如，在对TCP/IP协议进行补充之前，两个程序员必须在至少两种不同类型的计算机上构建软件并进行测试。另外，他们还必须证明这两个软件能够进行通信。

在INTEROP 92会议的主题致辞中，David Clark[8]将TCP/IP协议和Internet发展方式的特征描述为：大体上一致认同的，并且实际可运行的代码（rough consensus and working code），这句话阐明了一个简单的思想：尽管TCP/IP的许多部分得到多数研究人员认同，但是任何想法在

8 David Clark博士在1983年至1989年间供职于Internet架构委员会。

被实现和证明之前都是不被广泛接受的。

为了强调实用性和互操作性，Internet 项目的研究人员要求对方设计的软件能容错或处理不符合要求的包，他们致力于设计具有一定预测能力的软件，以发现软件在其它计算机上运行时可能出现的错误或缺陷，并尽力不违反 Internet 规范。

实现和测试工作总是在 TCP/IP 标准产生之前进行的，编写和测试程序经常能够发现设计中含混不清的地方和漏洞，并迫使设计者及早地纠正错误。因此，尽管 TCP/IP 协议非常复杂，但它的标准文档基本上没有错误。

20.7　成功的法则

许多接触 Internet 项目和 TCP/IP 技术的成功的人常常会问："我们能从中学到什么？"许多人问得更直接："我怎样能在自己公司的研究项目中也实现这样的成功？"

即使从行业内部人士的角度来看，这个问题也很难回答，因为这个项目是许多人在许多年的共同努力下完成的。下面是一些重点：

- TCP/IP 协议软件和 Internet 是由杰出的、有献身精神的人们设计的。
- Internet 是激励和挑战研究人员的一个梦想。
- 在没有短期经济效益的情况下也允许研究人员进行实验。事实上，Internet 研究经常使用的具有创新性的技术往往比现有的技术成本高。
- 研究人员并不是梦想迷造一个能够解决所有问题的系统，而是一个能够高效运行的系统。
- 研究人员在采纳一项标准之前，早已在实践中检验 Internet 的每个部分都运行良好。
- Internet 技术解决了如何连接多个网络这个重要的实践性问题，这是由于没有一个网络能在任何应用中都是理想的。

20.8　小结

Internet 是一项不可思议的技术成就。尽管认真的设计和对细节的关注对其成功有一定的帮助，但是研究人员对建立一个实际可运行系统的普遍认同更能帮助他们验证想法和排除错误。

第四部分　可以使用的 Internet 服务

人们如何使用 Internet

当前可以使用的 Internet 服务示例
以及一些令人感兴趣的
应用程序的介绍

第21章 电子邮件

21.1 引言

本章开始讨论Internet提供的服务实例，并介绍在Internet上应用最为广泛的服务之一——电子邮件服务。后面各章将介绍其他各项服务，对于每一项服务，首先描述其基本功能，并列举一项典型应用；然后描述其底层运作机制；最后，总结该服务的意义。

21.2 功能描述

电子邮件最初的设计目的是使两个用户能够通过计算机进行通信。世界上第一个电子邮件软件仅提供了一个基本功能：允许一个用户使用一台计算机键入一条消息，并通过Internet将这条消息发送给使用另一台计算机的用户。

目前的电子邮件系统提供的服务支持复杂的通信和交互。例如，电子邮件可以用于：

- 向多个收件人发送同一条消息。
- 发送包含文本、音频、视频、图片的消息。
- 向使用计算机程序进行应答的用户发送消息。
- 利用计算机程序对每一条收到的消息进行应答。

为了评价电子邮件服务的性能及意义，必须首先了解该服务的一些基本情况。以下几节将说明从用户的角度来看，电子邮件是怎样的服务。之后几节介绍电子邮件系统是如何工作的，并讨论电子邮件服务的影响。

21.3 全世界的福音

计算机网络的早期研究者认为，网络能够为人们提供一种将电信速度和邮件永久性结合起来的通信方式。计算机几乎能瞬时地将简短的消息或大的文件通过Internet进行传递。设计者们将这种新的通信方式称为电子邮件（electronic mail，e-mail）。电子邮件服务是一项在Internet上非常流行的服务。

21.4 电子邮件用户都拥有一个邮箱

为了接收电子邮件，用户必须拥有一个邮箱（mailbox），即磁盘上若干块存储空间，用来保存接收到的电子邮件消息，直到用户读取才能删除邮件。另外，邮箱所在的计算机必须运行电子邮件软件，当某消息到来时，电子邮件软件自动将其存储到用户的邮箱里。与邮局信箱一样，电子邮件邮箱也是私有的：任何人都可以向一个邮箱发送消息，但是只有邮箱的拥有者才能查看邮箱内容或取出邮件。

与邮局信箱一样，每个电子邮件邮箱都有一个邮箱地址（mailbox address）。要向另一个用户发送电子邮件，必须知道收件人的邮箱地址。总之：

> 使用电子邮件服务的用户必须拥有一个邮箱，该邮箱通过一个唯一的地址进行识别。只要知道对方的邮箱地址，任何用户都能通过Internet向另一个用户

发送邮件；只有邮箱的拥有者才能查看邮箱内容或取出邮件。

21.5 发送电子邮件消息

要想通过 Internet 发送电子邮件，用户必须在本地计算机上运行电子邮件应用程序。该应用程序与文字处理程序类似，它允许用户编辑消息，并通过给出邮箱地址以指定收件人。当用户完成输入消息和添加附件后，电子邮件软件就通过 Internet 将其发往收件人的邮箱。

21.6 电子邮件到达时的通知

可以通过配置系统软件，使得当有电子邮件消息到达时邮箱能及时地通知收件人。有些计算机显示出一条文本消息或者在用户的显示器上高亮显示一个小图标（例如，邮箱中的小邮件图案）；另外一些计算机发出某种声音或者播放一段示制好的消息；还有些计算机等到用户完成当前的应用程序后才通知用户。大多数系统允许用户选择取消邮件通知，这样用户必须定期查看邮箱中是否有电子邮件到达。

21.7 读取电子邮件消息

当有电子邮件到达时，用户可以使用应用程序从邮箱中获取邮件。该应用程序可以让用户浏览每封邮件，并有选择地对它们进行回复。通常，当电子邮件应用程序启动后，它会显示出邮箱中的邮件列表，并对每封邮件都给出一行摘要信息，包括寄件人姓名、邮件到达时间和邮件主题。在查看了摘要信息后，用户可以选择阅览该列表中的任何一封邮件。当用户选定列表中的一封邮件，电子邮件系统就显示出该邮件的内容。在阅览完一封邮件后，用户可以选择如下操作：回复寄件人、将邮件保留在邮箱中以备以后查看、将邮件备份至文件夹[○]、或删除该邮件。

总之：

> 如果用户想要发送和接收电子邮件，其连入 Internet 的计算机必须安装相应的应用程序软件。电子邮件软件允许用户编辑、发送邮件和读取收到的邮件。用户可以对任何邮件作出回复。

21.8 使用浏览器发送和接收电子邮件

如前所述，用户需要相应的应用程序软件来阅读和发送电子邮件。早期的电子邮件系统要求用户安装一个单独的应用程序，也就是说，用户要安装一个只能处理电子邮件的单独的应用程序。

许多现代的电子邮件系统采用了另一种做法：与使用单独的应用程序不同，允许用户使用 Web 浏览器（第 23 章中将会介绍）来存取电子邮件。通常，Web 电子邮件要求用户输入服务提供商的 URL。当用户输入了登录 ID 和口令后，该服务提供商则允许用户通过点击相应的链接来打开电子邮件账户。要点是：

> 用户必须利用相应的软件来存取电子邮件。一些电子邮件系统使用特定的应用程序，其他的电子邮件系统则使用 Web 浏览器。

21.9 电子邮件类似于办公室间的备忘录

电子邮件的格式类似于普通的办公室间的备忘录。邮件头指明寄件人、收件人、发送邮件的

○ 用户可以保存邮件的副本，作为一次会话的永久记录。

日期和时间以及邮件主题。与普通的办公室习惯一致，邮件头的信息显示在不同的行中，分别以 From、To、Date 和 Subject 开头，后面跟着相应的信息。例如，假设 2007 年 4 月 1 日，Jane 向 Bob 发送一封开玩笑的邮件。如果 Jane 的电子邮件地址是 jane@company1.com，Bob 的电子邮件地址是 bob@company2.com，那么该邮件的格式[B]如下：

```
From: jane@company1.com
To: bob@company2.com
Date: 01 Apr 2007 11:20:23 EST
Subject: some sad news
Bob,

    I heard that they've decided to cut personnel and you will be the first one fired
    April Fools!

Jane
:-)
```

在邮件的最后一行，Jane 附了一个符号暗示收件人清楚这是在开玩笑。这一行包括三个字符：一个冒号（:），一个减号（-）和一个右括号（）。这三个符号被称为“微笑”（smiley），因为它们组合在一起成为一个侧倒的微笑的脸。

电子邮件的邮件头还可以包含上述例子中没有的项目。例如，可以包含标注为 Cc 的一行，其中列出接收该邮件的其他人的邮箱地址[C]。

21.10　电子邮件软件填充邮件头信息

通常，寄件人只需要提供邮件头中 *To* 和 *Subject* 行所需的信息，因为电子邮件软件会自动填充时间和寄件人的邮箱地址。在回复中，电子邮件软件则会自动创建整个邮件头。它将原来的邮件中 *From* 域的内容填入回复邮件的 To 域中，同时也将原来邮件中 *Subject* 域的内容复制到回复邮件的 *Subject* 域中。软件自动填充邮件头信息为用户提供了方便，同时也使伪造邮件难以实现。

实际上，电子邮件系统提供更多的邮件头的行以帮助确定发送方计算机，指明寄件人的全名，提供惟一的邮件标识符用于审计和记账，确定邮件的类型（例如，文本或图片）。因此，电子邮件的邮件头可能会包含很多行[D]。一个过长的邮件头会让收件人感到厌烦，收件人必定会跳过邮件头去看信息的正文。总之：

> 尽管大多数电子邮件的邮件头包含很多行，但是电子邮件软件能够自动生成其中多数的行。友好的电子邮件软件也会在显示邮件时隐藏不需要的邮件头行。

21.11　电子邮件服务如何工作

回忆一下，第 17 章中提到，计算机间的通信一般涉及两个程序——客户机和服务器间的交互。电子邮件系统也是按照客户机/服务器的方式工作的：两个程序协同工作，将邮件从寄件人的计算机发送到收件人的邮箱（这个传送过程需要两个程序是因为在一台计算机上运行的应用程序不能直接将数据存储在另一台计算机磁盘上的邮箱内）。当某用户发送一封邮件时，寄件人计算机上的程序成为客户机，它与收件人计算机上的电子邮件服务器程序联系，并向它传送该

[B] 许多电子邮件应用程序向用户隐藏了邮件传输的格式。

[C] 在正式的办公室备忘录中，Cc 表示副本备份（carbon copy）：该术语来自在一张纸上打字用复写纸为文件多个备份。

[D] 在准备这章时，笔者收到一封垃圾邮件，其邮件头包含了 376 行！

邮件的一个副本，服务器程序则将该邮件存入收件人的邮箱。图21-1说明了这个工作过程。

用户完成一封电子邮件的编辑后，客户机程序就自动开始执行。客户机程序根据收件人的电子邮件地址来决定连接哪一台远程计算机，并使用TCP协议将此邮件的一个副本通过Internet发送到收件人的服务器程序。当服务器程序接收到一封邮件时，将其存储在收件人的邮箱中，并通知收件人有邮件到来。

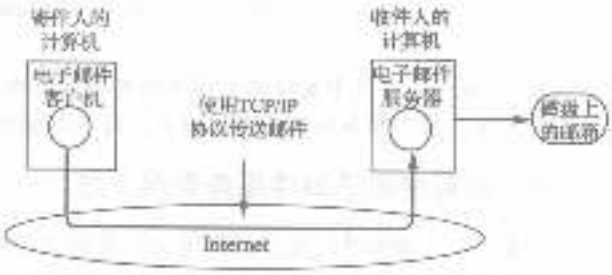


图21-1 在Internet上传送一封邮件需要两个程序，寄件人计算机上的客户机程序和收件人计算机上的服务器程序

客户机程序和服务器程序之间的交互是复杂的，因为双方的计算机或连接着它们的Internet随时都有可能发生故障（例如，某人突然关掉其中一台计算机）。为了确保电子邮件能够被可靠地传送，客户机程序必须在传送期间保留一份邮件副本。在服务器程序通知客户机程序该邮件已经接收并存储在磁盘上后，客户机程序才能删除该邮件副本。

21.12 在个人计算机上使用电子邮件

只有在计算机上运行着电子邮件服务器程序时才能接收电子邮件。在大型计算机上，系统管理员可以进行一定的配置，使系统首次启动时也启动服务器程序，并让其持续运行。服务器程序等待电子邮件的到来，将其存储在磁盘上相应的邮箱内，并继续等待下一封邮件的到来。

个人计算机经常关机或者断开与Internet的连接，当个人计算机未处于活动状态时，则无法接收到电子邮件。因此，大多数个人计算机并不直接接收电子邮件。用户通常在大型计算机上拥有一个邮箱，该大型计算机上的服务器程序时刻处于运行状态，接收到达的邮件并将其存入相应的邮箱中。例如，多数用户可以选择将他们的邮箱放在ISP的计算机上。为了存取电子邮件，用户则必须在个人计算机上运行相应的程序来连接ISP和获得他们的邮箱副本。

21.13 邮箱地址的格式

Internet上所使用的邮箱地址可能很长，使用起来较繁琐。邮箱地址由字符串组成，并被@符号分成两部分。前一部分指明了使用该邮箱用户的标识符，后一部分指明用户邮箱所在计算机的域名。在大多数计算机上，电子邮件系统使用用户的账号或注册号作为该用户的邮箱地址。例如，邮箱地址

jksmith@venus.engineering.somecompany.com

指明了在一台域名为venus.engineering.somecompany.com的计算机上有一个账号为*jksmith*的用户。计算机的域名是很难记住的。例如，德国的一台计算机的域名为：

14.informatik.rwth-aachen.de[⊖]

Purdue大学的一台计算机的域名为：

angwyshaunee.cs.purdue.edu

要想记住用户的标识符也是很困难的。实际上，用户标识符的格式取决于计算机系统以及管理员为分配标识符而制定的规则。一些系统使用用户姓名的首字、末字和中间名字的首字母

⊖ 在德国，与Internet连接的计算机的域名以de结尾，这是国际上代表德国的两个字母的缩写代码。

来构成用户的标识符，各部分之间用下划线隔开：

Jane_K_Smith@venus.engineering.somecompany.com

总之：

> Internet 上用于发送电子邮件的邮箱地址由字符串组成，并被圆点分成两部分。前一部分指明接受的用户，后一部分指明用户邮箱所在计算机的域名。

21.14 缩写使电子邮件服务变得友好

大多数电子邮件系统允许用户定义邮箱地址的缩写。缩写可以让一个公司为其各个部门建立更简短的名称，使公司的员工在寄信时不必键入冗长的后缀。例如，如果公司内所有的计算机都能将 eng 理解为工程技术部 venus.engineering.somecompany.com 的计算机的缩写，那么任何人从公司的计算机上发送邮件时就可以将邮箱地址写成：

Jane_K_Smith@eng

21.15 别名支持任意缩写

大多数商业电子邮件软件也支持电子邮件别名（email alias）功能，允许每个用户为他们经常使用的邮箱地址规定一组缩写。通常，别名机制要求用户准备一个简短的别名列表，由电子邮件软件将其存放在磁盘上。当电子邮件软件运行时，会寻找到用户的别名列表并使之生效。例如，假设一个用户经常向两个人发送电子邮件：*John Smith* 和 *Mary Doe*，他们的邮箱地址分别为：jksmith@computer1.somecompany.com 和 mary_doe@computer2.somecollege.edu。该用户可以为这两个邮箱地址定义两个别名：

```
john = jksmith@computer1.somecompany.com
mary = mary_doe@computer2.somecollege.edu
```

当用户编辑一封电子邮件时，可以将 *john* 或 *mary* 填入 *To* 域中。电子邮件软件会自动查询该用户的别名列表，并将缩写替换成相应的完整的邮箱地址。因此，尽管用户只是键入了别名，但发出的邮件中仍然包含了完整的邮箱地址。总之：

> 大多数商业电子邮件软件允许每个用户为他们经常使用的邮箱地址定义一组别名。如果用户在指定收件人时填入了别名，电子邮件软件会自动用完整的邮箱地址替代别名。

21.16 由一个计算机系统内的所有用户共享的别名

由于个人计算机系统是由单独的用户使用的，所以这种的系统只需要一组别名。但是，由许多用户共享的电子邮件系统则需要更加复杂的机制。除了每个用户都拥有一组私有别名外，共享的电子邮件系统一般允许系统管理员为所有用户定义一组共享的别名。当某用户指定一个收件人时，电子邮件软件首先查询该用户的私有别名列表，以确定是否有相应的别名存在。如果私有别名列表中没有该别名，那么就查询系统别名列表，以确定是否有相应的别名存在。

有了一组系统级的邮件地址别名后，系统的所有用户就可以共享这些别名。例如，假设系统管理员规定所有关问题报告都要发送给域名为

computer2.somewhere.com

的计算机上的 william 邮箱。为了方便用户，系统管理员可以选择一个易于记忆的缩写并创建一个系统级的邮件别名。如果系统管理员选择了缩写 trouble，那么如下的别名就可以加入系统邮件别名列表中：

```
trouble = william@computer2.somewhere.com
```

当该计算机上的任何用户想要报告一个问题时，他们可以将电子邮件发送给 trouble。电子邮件系统会查询别名列表中相应的目的地址，并将该电子邮件发送到邮箱 william@computer2.somewhere.com。

21.17 向多个收件人发送邮件

尽管电子邮件服务最初是为两个人的通信而设计的，但是电子邮件系统允许一个用户向多个收件人发送同一封邮件。这就需要寄件人在邮件头的 To 域中指定多个邮箱地址，系统会向每个收件人发送该邮件的一个副本。例如下面这封邮件：

```
To: bob@corp2.com, jim@corp3.com, susan@corp2.com
From: jane@company1.com
Date: 01 Apr 2006 12:34:02 EST
Subject: some sad news

Folks,
    I heard that your corporation is about to cut personnel and that you will be among the first ones
fired.
    April fools!

Jane

:-)
```

邮件头的 To 域中指定了三个收件人，其中两个人在 corp2，另一个人在 corp3。这三个人都将收到这封邮件的一个副本。

21.18 邮件列表：指定多个收件人的别名

电子邮件服务最有用的特性之一是它的别名机制有一个简单扩展：用一个别名指定多个收件人。当电子邮件系统查询其别名并找到多个收件人时，会向其中每个收件人都发送一份邮件副本。

非正式地，人们把指定多个收件人的别名称为邮件列表（mailing list）。例如，下面的定义创建了一个名为 *friends* 的邮件列表，其中包含 3 个电子邮件地址：

$$friends = bob@corp2.com,\ jim@corp3.com,\ susan@corp2.com$$

一旦创建了这个别名，任何发往 *friends* 的邮件都将会发送给这 3 个收件人。邮件列表概念可以简单地叙述为：

> 邮件列表是指定多个收件人的电子邮件别名；当有邮件发往该别名时，电子邮件系统会向该列表中的每个收件人发送一份邮件副本。

21.19 公共邮件列表和邮件检查程序

系统管理员可以创建公共邮件列表（public mailing list）。公共邮件列表允许与 Internet 相连的任何计算机上的用户都可以向该列表中的收件人发送邮件。例如，假设一台域名为 comp1.somewhere.com 的计算机上有一个名为 sales 的公共邮件列表[①]。要想向该表中的每个人发送邮件，就需要将邮件发送到地址：sales@comp1.somewhere.com。该邮件可以在任何计算机上生成。当该邮件到达目的计算机 comp1.somewhere.com 时，邮件检查程序（mail exploder）会检查该邮件的 To 域，找到列表名 sales 并将其扩展，然后向列表中的每个收件人发送一份邮件副本。图 21-2 说明了这个过程。

① 邮件列表的名字可以是有意义的或无意义的：*sales* 可以表示从事销售工作的一组人员，或者表示接受以销售服务为主题名，或者仅为了与其他列表中的成员进行区别。

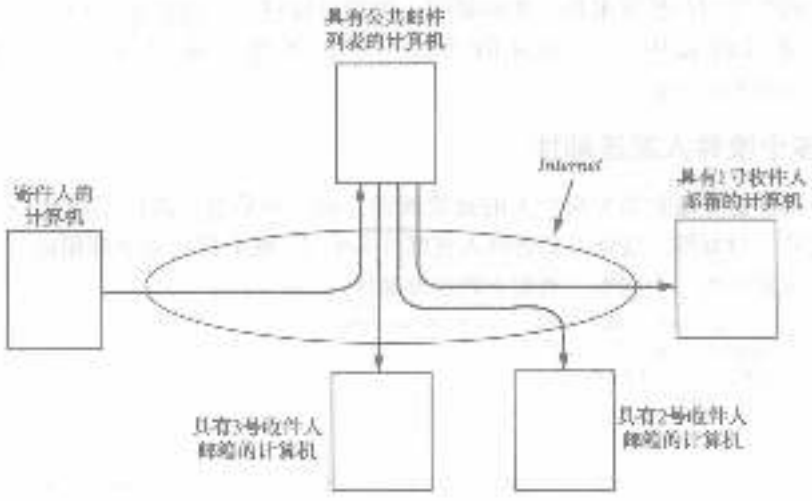


图 21.2　向一个包含 3 个收件人的公共邮件列表发送一封邮件所经过的路径，邮件检查程序接收该邮件，并转发至该列表中的各收件人

21.20　与非 Internet 网站交换电子邮件

Internet 上的电子邮件非常有用，这是因为它能被发送至其他网络。特别地，可以将电子邮件发送至具有通信能力的移动电话或呼叫器。同样地，也可以在移动电话上生成电子邮件，并将其发送至任意的 Internet 邮件地址。

为了发送电子邮件，需要一台中间计算机将 Internet 和移动电话的邮件系统连接起来。从本质上说，这台中间计算机能够说两种语言：它从 Internet 或移动电话系统接收电子邮件，并根据一组规则将其发送出去。通常，用户只有在已经付费或者与服务提供商签订了收费协议的情况下才能发送这样的邮件。

21.21　利用电子邮件访问其他服务

因为可以通过编程使计算机自动地响应电子邮件，所以任何与 Internet 相连的计算机系统都能利用电子邮件提供访问服务。更重要的是，大多数计算机上的电子邮件软件都能够方便地提供访问服务。

例如，假设某台与 Internet 相连的计算机上有数据库软件正在运行，任何登录到该计算机上的用户都能访问该数据库软件以查询数据库。但对于这两台计算机来说，尽管它们都有 TCP/IP 协议软件，都与 Internet 相连，但一个远程用户却不一定能访问另一台计算机的数据库，除非这两台计算机有更多的软件：该用户的计算机需要有数据库客户机程序，而被访问的数据库所在的计算机需要有数据库服务器程序。

上述客户机程序和服务器程序安装起来很麻烦，购买费用也很高。但是，如果服务器程序建立起来，在其上添加软件来提供利用电子邮件的访问服务却比较容易。为了实现这个功能，系统管理员需要编写一个计算机程序来对收到的电子邮件进行响应，将邮件传送给数据库软件，并将查询结果通过邮件发送回寄件人。这听起来很复杂，但是专家却能用若干行代码快速地创建这样的程序。其思想是：

> 因为计算机程序能够响应并回复电子邮件，所以电子邮件可以用于对多种远程服务进行访问。

21.22 速度、可靠性和期望

Internet的电子邮件软件是高效可靠的。一封电子邮件通常只需几秒钟就可以抵达目的地。如果目的计算机关机或者临时与Internet断开了连接，那么电子邮件软件会周期性地自动重传该邮件。此外，电子邮件服务比最好的邮政系统都要更加可靠：邮件几乎不会丢失，因为在邮件的分类和传送过程中都没有人的参与。如果一封电子邮件在一定时间内（通常是3天）仍没有到达，电子邮件软件会自动通知寄件人。

由于电子邮件能够提供高质量的服务，所以大多数Internet用户都认为它是可靠、高速的信息传递机制。用户将重要的信息通过电子邮件传送，而不使用其他的通信工具。更特别地，用户不会通过邮政信箱重发信件，也不会打电话确定电子邮件是否到达。

此外，由于邮件能够快速地到达，所以许多Internet用户认为电子邮件是一种即时通信机制。比起邮政服务来，它更像是电话服务。如果用户在收到一封电子邮件后快速地进行回复，那么当回复传回时，寄件人可能还在计算机上工作。有时，两个用户会交换一系列简短的电子邮件来进行会话。这样的通信一般出现在关系要好的朋友之间，并且在几次信件往来后结束。

但是，并不是所有人都认为电子邮件是重要的通信机制。有些人几小时、几天甚至几星期都不回复电子邮件。一部分人只是不想被邮件打扰，另一部分人则是忘了查看电子邮件。特别地，那些不使用计算机进行工作的人不能及时接到电子邮件到达的通知；他们要想查看电子邮件，还必须中断正常的工作并找到一台能使用的计算机。此外，由于许多家庭并不把计算机一直接入Internet，所以一些人无法方便地在家里查阅电子邮件。用户对电子邮件的认同度以及习惯的差异，会影响这项服务的使用，因为用户知道并不是所有人都会快速地回复电子邮件。

21.23 电子邮件服务的影响和意义

电子邮件服务的影响是如此显著，以至于很难对它作出评价。对许多Internet用户来说，电子邮件是必不可少的，实际上，电子邮件已经取代了邮政通信，成为这部分用户的主要通信手段！电子邮件改变了他们的日常生活，他们使用电子邮件与朋友、同事、雇员、顾客以及家庭成员进行通信。

要评判电子邮件服务，首先必须亲自使用它。通过大量使用，就会了解到电子邮件服务有如下优点：

电子邮件服务能够高速地传递信息，并且允许收件人自行选择何时作出回复，所以它既拥有即时通信的优点，还保证了用户不被中途打断工作的自由。

- 由于邮件列表允许任意一组成员相互交换信息，所以它提供了一种方法使有相同兴趣的一组成员能够共同参与讨论。
- 由于大多数计算机网络都提供电子邮件服务，并能与Internet上的电子邮件服务进行交互，所以与其他Internet服务相比，会有更多的人使用电子邮件服务进行通信。
- 由于电子邮件中可以包含文本、图片以及语音信息，所以它可以被用于传送文档和录制好的音频信息。
- 由于计算机程序能够自动地响应和回复电子邮件，所以Internet上已开发了许多服务项目，使用户能够通过电子邮件提出请求和获得响应。

总之：

> 虽然电子邮件服务最初是为了实现两个用户间的文本通信而设计的，但是它已经得到了极大的扩展。它允许一组用户进行音频、图片的通信，以及用户与

计算机程序间的通信。因此电子邮件服务已经成为 Internet 上使用最为广泛的服务之一。

21.24 加入邮件列表的约定

想要将自己的名字加入公共邮件列表的用户必须发送电子邮件请求，但是，这个请求并不一定发送给该列表。大多数提供公共邮件列表的 Internet 网站都约定：用于申请加入或退出邮件列表的请求邮件需要发送给另一个邮件别名。

为了理解这点，假设一位名叫 Judy 的用户，她的邮箱地址为：judy@bluechip.com，她将知有人会向一个邮件列表：

ballroom@athena.mit.edu

发送舞蹈信息。如果 Judy 向该列表发送一封邮件，那么该列表上的每个收件人都会收到一份邮件副本。为了加入某列表，要发送一封电子邮件至另一个地址，即在列表名后添加“-request”。因此，要加入上述舞蹈列表，Judy 需要发送一封电子邮件至：

ballroom-request@athena.mit.edu

这封邮件会被发送给维护该列表的人（或被发送给会回复更多说明事项的计算机程序）。当 Judy 加入了该列表后，她会收到所有发送到该列表的邮件的副本，并可以进行回复。

有些网站并不为请求设立别名。在这种情况下，用户可以尝试发送给邮件管理员（postmaster）。一般地，发送给邮件管理员的电子邮件会被传送给维护该电子邮件系统的人员。

第22章 电子公告板服务（新闻组）

22.1 引言

第21章介绍了Internet电子邮件服务，并说明了它是如何被用于交换信息的。本章介绍另一种Internet服务，它扩展了电子邮件服务的使用，允许用户加入一个或多个讨论组，并可以与组内其他成员共同参与讨论。电子公告板服务是Internet提供的最著名的服务之一，其上的讨论包含了十分多样的话题。本章将向用户介绍如何认识这项服务以及它的底层机制，并提供一些如何使用电子公告板和电子邮件服务的建议。

22.2 功能描述

电子公告板服务允许用户参加多个讨论组，各讨论组引对其个特定话题进行讨论。电子公告板服务允许用户：

- 选择一个或多个感兴趣的讨论组。
- 周期性地检查在其个讨论中是否有新的条目出现；若有，可以阅读其中一些或全部条目。
- 向讨论组发布一个条目以供其他人阅读。
- 发布一个条目，作为对其他某人所写条目的回复。

从概念上讲，Internet电子公告板服务在Internet中的作用正如普通的公告板在日常生活中的作用一样，它允许某人发布通知以供其他人阅读。实际上，Internet提供了对许多电子公告板的访问服务，它们都分别针对某一话题进行不断的讨论。所以，电子公告板服务有时也被称为计算机讨论组（computer discussion group）或计算机会议服务（computer conference service）。

为了能在许多用户中间进行有效的讨论，电子公告板服务结合了许多通信机制的特色。

- 与普通的公告板一样，电子公告板允许任何人发布消息以供其他人阅读。
- 与报纸一样，电子公告板服务将每条消息发布给众多订阅者。
- 与俱乐部或社会团体的通信一样，发布在某个电子公告板的消息是针对某一话题的。
- 与电子邮件服务一样，电子公告板服务能够快速地传送每条消息。
- 与社交聚会上的非正式讨论一样，电子公告板服务允许用户听取一段对话，提出问题，偶尔插入小的评论，或者发表长篇评论。

22.3 具有多样话题的电子公告板

Internet电子公告板包含非常多样的话题。例如，Internet上有关于科学、娱乐、政治、烹饪、架构、漫画书、科幻小说、诗歌、产品或服务、电影、股票价格、电视节目、流行音乐和计算机的电子公告板。尽管许多讨论一般人都能够理解，但是有些话题需要非常专业的知识，还有些话题只对在某个特定的公司工作或在某个特定的国家居住的人才有意义[1]。所以电子公告板可以被局限于某些计算机，某个组织或一小块地理区域，或被发布到世界各地的网站上。例如，加

---

[1] 除了文本信息，现代的新闻组还允许用户共享图片、音乐、电影和计算机软件。

拿大 Alberta 的一个包含政治讨论的公告板只能被发布到附近的网站上。

电子公告板多样性的一个原因是它们易于被创建。在某组织创建了一个电子公告板并开始讨论后，不可避免地会出现倾向于某些兴趣的话题的讨论。例如，假设某人在一个关于烹饪的讨论中提到某种品牌的厨具。如果其他人看到这条消息，并对该厨具发表他们的意见，那么就产生了一个新的讨论。如果这个关于厨具的讨论继续变得更加热门，那么参与者就可以创建一个新的电子公告板，并把关于厨具的讨论转移到新的电子公告板上来。分裂一个电子公告板就类似于人们在聚会上组成小型的讨论组，即可同时进行多个会话而不互相干涉。要点是：

> 由于很容易创建新的电子公告板，并且根据话题分聚电子公告板可以使电子公告板上的讨论更加集中，所以许多的电子公告板被创建出来。

22.4　网络新闻

最初，Internet 上的电子公告板服务被称为网络新闻（Network News，NetNews）。网络新闻系统使用术语新闻组（newsgroup）来表示每个单独的电子公告板（即每个讨论组），术语文章（article）表示发送给新闻组以供大家阅读的消息。每个提交给新闻组的消息与电子邮件消息的形式一致，可以短至几行或长至许多页。与电子邮件一样，每个文章的标题中都有一个 *From* 行来确定发送人。

最初网络新闻是早期的拨号网络的一部分，该网络使用拨号调制解调器在计算机之间进行拨号连接并传递信息。起初，网站通过拨号连接来传递网络新闻，并使用术语 USENET 来表示其计算机"网络"。由于 USENET 网站逐渐支持 Internet 访问，所以许多网站将其网络新闻的传递方式从拨号电话系统改为 Internet。尽管大多数参与网络新闻的网站现在都通过 Internet 接收信息，但是那些不支持 Internet 访问的网站仍然通过其他通信方式参与网络新闻，例如 BitNet 和拨号电话系统。

22.5　公告板的种类

最初的网络新闻系统为每个新闻组指定一个惟一的名字。新闻组名字的第一部分指明了该新闻组的类别，剩下的部分则指明了主题及该主题中的某个特定话题。例如，新闻组

rec. sport. baseball

包含了关于棒球的讨论。这个新闻组的类别是娱乐（recreation），它的话题是体育（sports），而主题则是棒球（baseball）[6]。

近来的电子公告板服务设计出了更多新的种类。例如，图 22-1 列出了当今 Internet 提供的一些主要的新闻组种类。

	主要种类
Arts and Entertainment	艺术和娱乐
Business and Finance	商业和金融
Computers and The Internet	计算机和 Internet
Cultures	文化
Family and Home	家庭
Games and Gaming	游戏和赌博
Health and Fitness	健康
Music	音乐
News and Current Events	新闻和时事

图 22-1　Internet 上使用的新闻组分类示例

---

(6) 尽管新闻组的名字与第 18 章介绍的计算机域名和第 15 章介绍的 IP 地址类似，但是它们之间并没有联系。

	主要种类
Politics and Government	政治和政府
Regions and Places	地域
Religions	宗教信仰
Romance and Relationships	人际关系
Schools and Education	学校和教育
Science and Technology	科学技术
Society and Humanities	人文社会
Sports and Recreation	体育和娱乐

图 22-1　（续）

22.6　获取网络新闻及阅读文章的软件

最初，用户必须运行单独的应用程序才能阅读和发布新的文章。但是，现代新闻组可以通过 Web 来访问。例如，以下网站：

www.groups.google.com

www.groups.yahoo.com

www.usenet.com[1]

每个网站都包含主要的新闻组种类；Google 还提供对 USENET 存档的访问。此外，其他两个网站提供对电子公告板服务的 Web 访问。

22.7　用户眼中的网络新闻

选择一个类别是使用新闻组的第一步。在大多数情况下，每个类别都被再分为一组话题。因此，当某用户选择了一个种类之后，系统会向该用户显示出该种类中的所有话题。例如，如果某用户选择了种类 Computers and The Internet，该用户将会看到以下话题：

计算机硬件

计算机软件

数据网络

Internet

计算机教育

计算机的社会影响

有些话题的范围太广所以被更细分成不同的子话题。例如，在计算机软件这个话题中，有子话题 Microsoft Windows、Linux 等。因此，用户在选择了某个话题后，会看到一个子话题列表。当该用户选定了某个子话题后，系统会列出一个新闻组列表并允许用户进行选择[2]。当然，当用户找到一个感兴趣的新闻组时，可以为这个新闻组保存一个引用；使用直接的引用可以使用户不必在以后的访问中从许多的种类、话题和子话题中寻找该新闻组。

22.8　查阅新文章

如果某用户加入了一个新闻组，如何才能知道是否有新文章出现呢？一种方法是人工不断地查看该新闻组，例如，用户可以每隔一天访问该新闻组以查看新文章。还有另一种方法，许多新的访问接口允许用户进行设置，在有新的文章出现时，可以发送一个提醒（alert）给用户。

---

[1] Usenet.com 是一项付费订阅服务。

[2] 一些新闻组是对所有用户开放的，另外一些新闻组要求用户注册成为会员后才能阅读文章。

总之：

> 参与网络新闻的用户可以选择自动接收提醒或者定期查看新闻组来阅读新文章。

22.9　阅读网络新闻

阅读网络新闻是很简单的。当用户选择了某个新闻组时，系统则会显示出该新闻组中的新文章列表。该表中的每一项都包含某篇新文章的摘要信息，指明该文章的作者（通常是一个电子邮件地址）和内容（一行文章主题）。要想阅读一篇文章，用户需要点击该文章来使文章内容在窗口中显示出来。

新闻文章的格式与电子邮件是一致的。即，每篇文章都包含两个部分：标题和正文。与电子邮件的邮件头一样，新闻文章的标题包含：From 行表示发布该文章的人的电子邮件地址，Date 行表示该文章发布的日期和时间，Subject 行列出该文章的话题。也可以包含其他行，如 Distribution 行表示该文章的限制传播范围。例如，以下一篇文章：

```
Article 530 in rec.motorcycles.harley:

From: rick@ company1.somewhere.com
Newsgroups: rec.motorcycles.harley, soc.culture.japan
Subject: Harley Davidson in Japan
Date: 8 Dec 2007 15: 35: 56
Organization: Somewhere Incorporated, Fremont, California
Distribution: world

Folks,
    A friend of mine told me that Harley Davidson went to Japan to sell motorcycles in 1953. He says
that H-D started the Japanese motorcycle industry, and that the Japanese bikes eventually outsold
Harley-Davidson. Is there any truth to the rumor? Does anyone know of a book that tells the story?
```

22.10　提交文章

向网络新闻组提交文章就像发送电子邮件一样容易。用户编辑一条消息，并请求将该消息发布在一个或多个新闻组中。如果一篇文章发布在多个新闻组中，标题中有一行会逐个显示这些新闻组的名字。

除了编辑消息外，用户还可以对现有的文章进行回复。与电子邮件一样，新闻阅读器会从用户所回复的文章中提取必要的信息来形成新的回复文章的标题。用户则可以编辑该回复文章然后进行发布。

22.11　受限新闻组

有些新闻组允许所有人提交文章。在这些新闻组中，提交文章的人可以是新成员，也可以是专家。更重要的是，其中有对某一话题感兴趣的认真学者，也有仅仅是开玩笑的人。

为了限制对新闻组的破坏及弥补各成员背景的差异，有些新闻组是受限的（moderated）。从本质上说，管理员可以预览所有提交给该新闻组的文章。管理员会决定一篇文章是否适合发布，并提交该文章、编辑该文章、选择并提交该文章的一部分，或者对几篇文章进行总结。如果管理员能够做好这些工作，那么相应的新闻组中就不会出现不相关的、误导性的或者破坏性的言论。

22.12　网络新闻的规模

网络新闻是很受欢迎的。2006 年，某商业网站公布有超过 120 000 个不同的新闻组，每天的

文章数超过两百万。此外，许多公司还有内部新闻组，只对公司雇员开放。

要得到新闻组的准确规模是不可能的，原因有两个。首先，除了一个全球范围的网络新闻服务外，还有许多服务提供商提供电子公告板服务。其次，由于 Internet 已经覆盖了全球，其他国家的组织还创建了局限于某国家或地区的新闻组。尽管新闻组的总规模无法估计，我们还是可以得出以下结论：

> 由于网络新闻拥有成千上万的新闻组，并且每天都会接受新的文章，所以没有人能加入所有的新闻组。

22.13　新闻组和邮件列表的影响及意义

电子公告板和公共邮件列表的影响是很难估计的。Internet 技术使数百万的人参与电子讨论。正由于有这么多人的参与，所以很难想像出现的话题会有多么千奇百怪。更重要的是，新的讨论不断出现，讨论某个话题的一群人可能很快会对另一个话题感兴趣。许多人抱怨公共邮件列表和新闻组更新得太快，以至于他们无法参与所有他们感兴趣的讨论。

Internet 电子邮件列表和电子公告板服务带来了一个值得关注的社会影响：它们使世界各地的人们有机会交流各自的观点。过去，这样的交流局限于居住得接近，能够面对面交谈的人们（例如，邻居们在后院交谈），或者一些在报纸和杂志上发表观点、专栏或社论的人。值得关注的是，由于 Internet 超越了地理和政治的界限，所以讨论能够扩展到来自许多不同国家的人中去。

Internet 上的社区的组织形式与实际地理区域中的社会团体完全一样。兴趣相同的人会相互发现。有时，他们会找到与自己观点一致的人。但讨论发生分歧时，新闻组或邮件列表就会分裂成两个，使各组成员都拥有相同的观点。

当然，使用网络新闻或电子邮件进行交流与普通的交流是不同的。首先，由于大多数电子邮件备忘录或文章都是文本格式的，所以就不能通过声音或姿势来表达情感。其次，由于网络系统将每条消息或文章传递给许多人，所以几乎任何观点都能引起其他人的响应。第三，由于文章在几天后会过期，所以新加入的成员可能无法获得讨论的上下文或历史。

22.14　参与讨论的建议和约定

网络社区的成员需要遵守一定的规则来保持文明的社会交流，并能辨别文明的行为和不文明的行为。学习 Internet 交际的正确礼节是困难的，这有两个原因。首先，因为 Internet 跨越了许多不同的文化、经济背景和教育层次，所以很有可能正在交流的两个人拥有不同的背景。其次，因为 Internet 是一种新兴的通信媒体，所以有些人会误以为不需要规则约束。实际上，确实有一些用户并不遵守文明讨论的原则。

不同背景之间的差异通常在一些微妙的方面变得很明显。在一些文化中，人们会互相信任；而在另一些文化中，则允许分歧。同样的，不同的文化对同一地位、阶层或头衔会有不同的态度。事实上，其种头衔在一种文化中很重要，在另一种文化中可能根本没有意义。为了弥补这些差异，有一些简单的准则可供参考[①]：

- 在阅读电子邮件备忘录或新闻文章时，不要对作者作任何假设。作者的经验或在某个话题上的专业知识可能比你多，也可能比你少。简而言之，就是不要假定作者很专业或者很愚蠢。
- 要敏于对著名人士或知名权威发表的消息提出怀疑，因为标题信息可能被伪造，有些人

① 作者根据的搜索到了这个列表，Web 上它有类似的建议列表，称为网络礼节（Netiquette）。

会以此为乐。

- 当编辑一条消息时，要考虑到会有很多不同背景的人阅读它。要选择能准确表达你的观点的词语。如果你有可用的资料的话（例如，引用书或杂志中的文章），要为你的观点提供论据。
- 在任何社会交流中都要遵守一定的约束。例如，在对一些具有煽动性的、过分的观点作出回复前，要仔细地考虑一下。
- 不要进行人身攻击，特别是对那些回复你所写文章的人。要记住作者并不认识你，也可能并不看重你的头衔或地位。
- 使用表示微笑的符号：），来表示你使用的是幽默的表达方式。
- 许多 Internet 用户都遵循一定的大、小写的惯例：全大写的词语表示惊叹。
- 如果你是初学者，请从寻求帮助开始。特别地，一些公共邮件列表和新闻组都维护一个称为常见问题（Frequently Asked Questions，FAQ）的摘要文件；另外一些则会保存过去讨论话题的摘要。你可以提交一条消息来询问是否有在线的 FAQ 或讨论摘要。

22.15　小结

Internet 上的电子公告板服务允许用户对某一特定话题发布文章。成千上万的新闻组中包含多种多样的话题：教育、爱好、政治、科学、娱乐、就业机会等。尽管许多的新闻组在世界范围内发布，有些却被限制于特定组织、城市、国家或大陆。

许多用户通过浏览器来访问新闻组。访问接口允许用户查找存在的新闻组，阅读其新闻组中的文章，或发布新的文章。新闻组和电子邮件使用相同的格式：标题指明发送人，正文是文章内容。

第 23 章　浏览万维网

23.1　引言

第 22 章讨论了用于在用户之间传递信息的基本 Internet 服务。本章将介绍一种动态服务，它允许用户浏览存储在远程计算机上的信息。本章会描述浏览服务是如何工作的，并解释使万维网广为流行的超媒体的概念。第 24 章将会介绍浏览服务的更多细节。

23.2　功能描述

信息浏览服务（information browsing service）使用户能够方便地获取并显示存储在远程计算机上的信息。大多数信息浏览服务都采用交互式的操作方式：用户提出请求，浏览系统则返回相应项目的副本来作为响应。如果返回的项目中所包含的信息的形式适合人们阅读，那么浏览器会自动在用户的屏幕上显示该信息。但是，我们也会看到，浏览服务还能完成许多其他任务。大体上，浏览器可以：

- 从远程计算机上的各种信息源获取文本信息、录音、图像或视频。
- 自动显示获取的信息。
- 将获取的信息存储在磁盘上。
- 将获取的信息打印出来。
- 通过某一文档中的引用来链接到可能存放在其他计算机上的相关文档。

23.3　浏览与信息检索

从广义上讲，信息检索服务与浏览服务都给用户提供了从远程计算机上获取信息的机制。早期的 Internet 服务使用检索方式：用户指定要检索的项目，检索服务则提供相应的项目副本，并将其存储在用户的计算机上。信息检索服务与浏览服务最主要的不同之处在于它们向用户提交信息的方式不同。尤其是，虽然信息检索服务允许用户获取远程文件的副本，但是它却并不向用户提供该文件的格式或内容等相关情况。要点是：

> 尽管信息检索服务能有效地传输数据，但是它却不能向用户显示由文件的内容。

与此相反，浏览服务能够显示远程计算机上的信息，并且不需要用户记住该信息所在的计算机、文件或目录的名字。事实上，大多数的浏览服务都使用简单的点击式界面，以使其易于使用。其思想可以归纳如下：

> 浏览服务使用户不必知道每个文件的名字就能浏览远程计算机上的信息。在获得了远程计算机上的一个文件副本后，浏览服务会自动显示其内容，并允许用户选择其他相关的文档。

23.4　早期的浏览服务采用菜单驱动方式

迄今为止，已经开发出几种用于 Internet 上的浏览服务。gopher[1] 是最早推出的浏览服务之一。由于它是最早普及的浏览服务之一，所以许多早期的 Internet 网站都运行着 gopher 服务器。因此，gopher 能提供对大量信息的访问。由于 gopher 的界面直观、易用，所以它在初学者和非专业的 Internet 用户中尤其流行，用户不需要学习命令就能够使用 gopher。

早期的 Internet 服务普遍采用菜单驱动的方式来进行信息浏览。这类服务的界面上显示一个可由用户进行选择的菜单，并且允许用户在其上选择感兴趣的项目。由于早期的服务主要针对文本信息，所以每个菜单项都包含简短的、自说明的短语。为了使系统易于使用，菜单中项目的个数一般会受到限制，以使每个菜单能够恰好与用户屏幕相匹配。在实际系统中，菜单上菜单项的个数一般限制于 12 个左右。

菜单驱动系统的不断流行是因为它使用户能方便地从菜单上的一个菜单项转移到另一个菜单项上去。从理论上讲，菜单上的每个菜单项都包含一个对附加信息的隐式引用。这些附加信息可以是一个数据文件或另一个菜单。

浏览器软件在显示出菜单后，会等待用户阅读该菜单并选择其中一项。如果所选项目的隐式引用指向一个数据文件，那么浏览器软件会获取该文件的一个副本并显示其内容。如果所选项目的隐式引用指向另一个菜单，那么浏览器软件会获取该菜单，将其显示出来，并等待用户作出另一个选择。要点是：

> 菜单驱动浏览系统显示一个可供用户选择的菜单，用户通过选择菜单中的项目来向浏览服务请求检索文件中的信息或检索另一个菜单。

23.5　菜单项可以指向另一台计算机

Internet 浏览系统能够跨越多计算机系统，因此它具有强大的功能：一台计算机上的一个菜单项可以指向另一台计算机上的一个菜单。由于菜单项的引用是隐式的，所以用户不必做任何特殊的操作就能链接到另一台计算机上。事实上，用户可以阅览一系列菜单，而不必在意这些菜单是来自一台计算机还是多台计算机。更重要的是，由于浏览器软件能够很快地获取一个新菜单，所以用户不会感觉到很长的延迟。从用户的角度来看：

> 浏览系统完全隐藏了计算机间的界线，并且使许多计算机上的信息成为一个集成系统的一部分。浏览器可以从一台计算机链接到另一台计算机，而用户并不需要知道或在意正在访问的是哪台计算机。

23.6　浏览器如何工作

与其他 Internet 服务一样，浏览服务也使用第 17 章中描述的客户机/服务器方式进行交互。图 23.1 说明了浏览服务的工作原理。

如图 23.1 所示，浏览器程序运行在用户的本地计算机上，用户与浏览器程序进行交互来控制信息的选择与显示。例如，用户指定一个初始菜单，然后选择一个菜单项让浏览器进行相应的工作。浏览器则作为客户机来获取信息。在每次用户选定了一个菜单项后，浏览器提取出与该菜单项有关的隐式引用信息，并使用该信息来确定联系哪个远程服务器以及获取该服务器上的哪

---

[1] 之所以选择这个名字，是因为该系统是在 Minnesota 大学设计开发的，那里是金色鼠（Golden gopher）的发源地；这个名字还是一个双关语，因为该系统以前“跑腿”（go for）信息为目的来设计的。

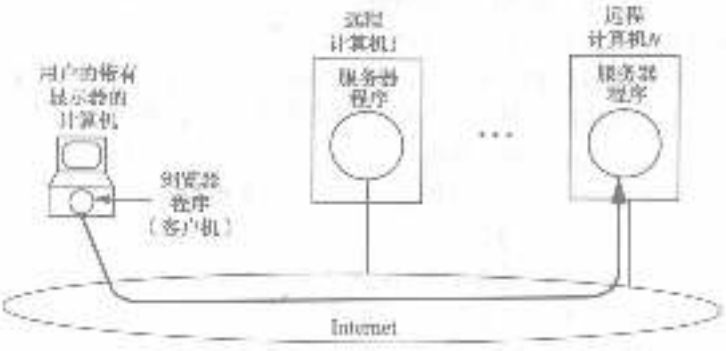


图 23-1　浏览服务使用客户机/服务器方式进行交互。多服务器持续运行；浏览器代为客户机每次只与一个服务器联系

个项目。新选择的项目可能与当前菜单来自同一台计算机或来自不同的计算机。浏览器每次只能与一个服务器联系：浏览器与服务器建立连接，获取所需要的信息，关闭该连接，并向用户显示相应的信息。为了使系统能有效地工作，服务器必须时刻保持运行状态，并且能向任何提出请求的客户机提供相应信息。

23.7　点击式界面的示例

为了理解信息是如何存储在简单的菜单项中的，考虑一下使用文本菜单的点击式（point-and-click）交互方式，即，相应的软件允许用户仅仅通过使用鼠标就能遍历菜单，而不需要使用键盘。例如，鼠标可用于从菜单中选择一个项目，返回之前的菜单，或退出该应用程序。

图 23-2 是点击式界面的一个例子。屏幕被分成了三个主要区域。中间的区域包含了最重要的信息：菜单的标题和用户可选择的项目列表。本例中含有 13 个菜单项，每个菜单项分别对应于另一个菜单、一个数据文件或一个计算机程序。要选择其中某一个菜单项，用户需要移动鼠标，使光标（图中以一个箭头显示）位于要选择的项目上。当用户点击了鼠标按钮后，浏览器就会高亮显示光标下的项目并获取所选择的项目。例如，图 23-2 给出了用户将光标移动到第 8 个项目后的显示效果。

图 23-2　信息被组织成菜单的浏览系统示例，光标是一个箭头，并随着用户鼠标的移动而移动

23.8　菜单项与文本相结合

早期的浏览系统将菜单与数据文件相分离，显示器上或者显示用户可选的项目菜单，或者显示用户可读的文档。从本质上说，用户需要在菜单中进行浏览，直到发现感兴趣的项目。当用户选择了某个项目后，菜单会从

屏幕上消失，取而代之的是相应的文档。同样地，当用户阅览完其文档并选择重新浏览其他信息后，最近一次浏览的菜单会显示出来，以代替上一次阅览的文档。

浏览服务将菜单与其他信息相分离，可以易于分辨所显示内容的类型：屏幕上显示的是菜单还是其他信息总是不会混淆的。但是，这样的浏览服务通常会将每个菜单项的内容限制于单行的文本信息，因此，由于菜单项中不能提供更多的详细信息，所以它们可能会使人难以理解。例如，一个简单的菜单项中无法包含描述性文字、图片或其他信息。

如今的浏览服务采用了完全不同的方式。与分离开菜单和其他信息不同，它们直接将菜单项嵌入到文本中。嵌入式的菜单项更易于人们选择，因为它们周围的文字能够提供上下文关系。为了理解这一集成概念，考虑下面这个段落示例，这是它在支持嵌入式菜单项的浏览系统中的显示效果。

> *The New York Stock Exchange is a world renown center of business activity. Located on Wall Street in downtown New York City, the stock exchange allows stock brokers to buy or sell shares of stock. At the end of each day, the average price per share of stock as well as the total number of shares traded are computed and used as a measure of activity. Many newspapers list the closing stock prices each day in their business section.*

该例是对纽约证券交易所的一个简短描述，它向读者提供了一些诸如交易所所在位置等方面的情况。除了普通描述外，该段落中还有几个项目被加上了下划线以突出显示。这些突出显示的项目对应于读者可以选择的链接，就类似于面向菜单的浏览器中的菜单项。

当浏览服务显示某个段落时，用户可以选择阅读所有信息，也可以只浏览突出显示的项目。已经了解该主题的基本信息并且需要寻找详细信息的用户可以只查看突出显示的项目；当用户不清楚某个项目的含义时，可以阅读该项目周围的文本来作为参考。

由于突出显示的项目能够吸引读者的注意，所以查看嵌入式的项目能够像浏览菜单一样方便快捷。例如，用户可以浏览上面的示例段落，并在几秒钟内定位突出显示的项目，而不需要阅读整个文本。

与菜单驱动的浏览系统一样，每个可选择项目都含有一个指向另一个项目的隐式引用。当用户选择了一个突出显示的项目后，浏览器会作出一定的响应：转到相应的隐式链接，获取该项目的一个副本，并显示出新的信息。用户可以继续在新的显示内容中选择其他链接以获取更多的信息。例如，如果用户在上述段落中选择了项目 average price per share，那么浏览器会转到相应的链接，并打开描述股票平均价格的文档。如果用户选择了项目 Wall Street，那么浏览器会显示有关纽约市华尔街的详细信息。

23.9　集成式链接的重要性

尽管上面的示例非常短，但是它仍然说明了链接是如何嵌入到其他信息中的。集成式系统从两个方面帮助用户：

- 首先，同时显示信息和可选择链接能够更详细地解释一个主题，并使人们易于理解突出显示的项目。例如，项目周围的上下文通常能够帮助说明其含义并消除歧义。
- 其次，在信息中嵌入链接有助于用户对所遇项目进行研究。浏览服务还能进一步记住用户在选择某个项目时所处的位置，并在用户阅览完该项目后返回到这一位置。

使用集成式浏览系统就像读书一样。读者在遇到不认识的单词或短语时会停止阅读，并去查询字典。如果字典的定义中仍有读者不理解的术语，读者还可以继续查询。最后，读者在掌握

了所有相关的术语后会回到书中继续阅读。在信息中嵌入可选择链接的信息浏览服务就支持这样的功能：用户在查看某文档时，可以停下来去查询某个嵌入式链接的信息，然后再返回到该文档中继续阅览。

总之：

> 如今的浏览服务将可选择链接嵌入到其他信息中，消除了歧义，并且使用户更易于理解每个项目。将链接与其他信息结合起来有助于用户对感兴趣的项目进行研究。

23.10 在文本中嵌入的链接称为超文本

在文本中嵌入可选项目的概念称为超文本（hypertext）。超文本并非起源于Internet浏览服务，它已经被用于普通的计算机程序很多年了。例如，计算机制造商使用超文本系统来显示计算机及其上程序的使用说明文档。用户首先要求系统显示计算机的说明文档，用户可以像阅读普通文章一样来阅读该文档，也可以在其上选择任何项目。当用户选择了一个单词或短语时，该超文本系统就查找该项目，并显示与其相关的信息。

为了理解存储在计算机系统中的超文本文档，想像一下桌面上有一叠纸质文档，每个文档中对其他文档的引用的项目都被突出显示了。以图23.3中所示的6个文档为例，每个文档到其他文档的引用都用箭头表示。

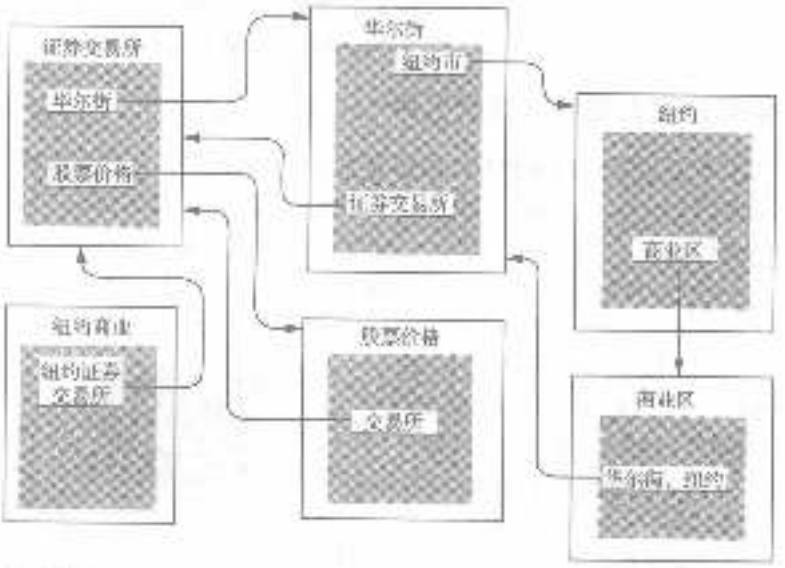


图23-3 6个文档示例，忽略文档内容，只保留文档标题和对其他文档的引用。在单词或短语与它们所引用的文档之间画出箭头表示

在本例中，标题为“华尔街”的文档包含两个引用，一个是“纽约市”，另一个是“证券交易所”。引用“纽约市”指向标题为“纽约”的文档，该文档中包含一个对“商业区”的引用。

当把超文本看作是一组文档时，有以下两点需要说明：

- 突出显示的短语并不需要与其所引用的文档的标题相同。例如，标题为“股票价格”的文档中包含突出显示的短语“交易所”，但是该短语的隐式引用所指向的文档标题为“证券交易所”。超文本文档的设计者指定需要被突出显示的短语以及每个短语所指向的文档的确切位置。

- 并非所有文档都包含相同数量的引用，某个特定文档中可以不包含引用、包含多个引用，或者仅包含几个引用。

当然，在使用超文本系统时，用户并不会看到各文档之间联系的图形说明。相反，屏幕上一次只能显示一个文档，用户只有在点击了某个突出显示的短语后，才能知道该短语指向哪个文档。

总之：

> 虽然超文本系统中的文档之间可以包含复杂的引用关系，但是用户却看不到这样的复杂性，因为用户一次只能阅览一个文档。

超文本的强大源于它将通用性和即时引用结合了起来：每个文档都能包含文本，或包含指向其他文档的嵌入式菜单项。用户可以选择阅读整个文档或仅查看突出显示的短语。无论何时，用户都能返回至上一个文档，或选择指向另一个文档的突出显示的短语。

23.11　多媒体

要想理解 Internet 浏览服务是如何扩展超文本的，就必须先了解现代计算机系统的工作能力。尽管早期的计算机只能以打字机字体显示文本，但是现代的计算机系统却拥有复杂的硬件用于显示图像、播放视频和声音。这些计算机使用图形硬件，并以比例间距来显示文本：为每个字符分配的空间与字符的宽度成正比[⊖]。现代的计算机系统还能够显示多种彩色、几何图形和图表、静止或动态的图片，以及高质量的音频。由于能够播放音频和视频的计算机系统的输出可以支持多种媒体（multiple media），所以这样的计算机被称为多媒体（multimedia）计算机。

23.12　文本中可以嵌入视频和音频引用

多媒体计算机的出现，使访问多种形式的信息成为可能。例如，下面这段话：

> *Several city tours include visits to the New York Stock Exchange. During a typical working day, many people crowd onto the floor of the exchange. Visitors can hear the sounds of traders bidding and view the sights of trading activity recorded by cameras that look down on the scene.*

该段文字描述了纽约证券交易所的一天和参观者能所见所闻。与之前的例子一样，可选择的短语会用下划线突出显示。但是，与之前的例子不同的是，一些突出显示的短语并非对应于文本信息。如"view the sights"和"hear the sounds"就指向了视频和音频信息。

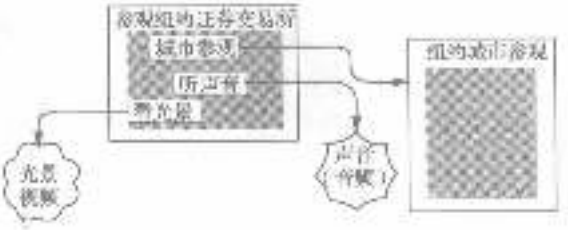


图 23-4　超媒体文档的逻辑结构。当用户选择"听声音"时，计算机会播放存储的音频。当用户选择"看光景"时，计算机会播放存储的视频

当超文本系统中含有对非文本信息的引用时，就称其为超媒体（hypermedia）系统。图 23.4 举例说明了一个含有对非文本信息引用的超媒体文档的逻辑结构。

在计算机中，非文本数据必须存储为数字形式。例如，计算机将声音存储成数字序列，与光盘或 MP3 播放器存储音乐的方式类似。计算机也使用数字存储图片和视频剪辑。当然，播放音频或视频的用户不会看到底层的数字序列。

⊖ 由于本书采用了比例文本进行排版，所以字母 i 比字母 w 所占的宽度要小。

总之：

> 超媒体系统可以在文档中嵌入对非文本信息的引用和对文本信息的引用。如果用户选择了对文本的引用，超媒体系统则会显示相应的文档；如果用户选择了非文本引用，超媒体系统则会播放相应的音频或显示相应的图像。

23.13 万维网

时下最流行的支持超媒体的Internet浏览服务，即万维网（World Wide Web，Web）[1]。这项服务是将多台计算机上的信息连接起来的一种机制。从本质上说，Web支持在一台计算机上的文档中引用其他计算机上的文本或非文本信息。例如，美国某台计算机上的万维网文档中可以包含对瑞士某台计算机上存储的视频的引用。

用户浏览万维网就像某人在某台单独的计算机上浏览超媒体一样。假设某个时候，用户的显示器上有一个含有突出显示的引用的文档，当用户选择某个项目后，系统会根据该引用，获取所指向的项目，并播放相应的声音或显示相应的文档。因此，用户不必知道信息的来源就可以浏览万维网。

尽管理解超媒体服务的工作原理很困难，但是它使所表达的信息可视化的能力却使它很容易被接受。因此，万维网是最流行的服务[2]。即：

> 除了包含文本信息外，Web文档中还可以包含声音、图像和视频。为了显示非文本信息，计算机必须配置多媒体硬件。

23.14 用于访问Web的浏览器软件

用于访问万维网的软件被称为Web浏览器（Web browser）。第一个Web浏览器是由美国国家超级计算机应用中心（NCSA）开发的，叫做Mosaic。该浏览器能够控制用户的显示器，并且允许用户使用鼠标来浏览超媒体文档。现在，Microsoft等公司都提供商业的浏览器，此外，Mozilla计划开发出了很受欢迎的开源浏览器。

当浏览器显示某文档时，其中的一些项目会链接至其他文档。对于链接指向的文本，浏览器在链接下方添加下划线以突出显示，并且使用与周围文本不同的颜色来显示它（例如，如果浏览器用黑色显示普通文本，那么突出显示的文本可以是蓝色的）。要想选择一个项目，用户需要移动鼠标，直到光标移动到要选择的项目上，然后用鼠标单击此项目。作为对某个选择操作的响应，浏览器会根据相应的隐式引用，获取指定的项目，并且将它显示给用户。如果被选择的链接指向一段录音，那么Web浏览器会播放该录音。如果被选择的链接指向一个图片或照片，那么Web浏览器会在用户的屏幕上显示该图像。如果被选择的链接指向一个视频剪辑，那么浏览器会获取该视频并像电视画面的播放一样显示它。

尽管大多数突出显示的项目是由文本组成的，但是Web文档中也允许突出显示非文本项目。例如，如果一个图像代表一个链接，那么该图像将整个被突出显示；点击该图像会使浏览器打开相应的隐式链接。事实上，一个图像的不同部分可能会代表不同的引用。因此，一个Web文档中可能会含有一幅图片，其中每个图片都包含指向另一个Web文档的链接。简而言之：

> Web浏览器是由一个程序构成的，该程序提供对万维网上超媒体文档的访问。

---

[1] 最初开发万维网的目的是支持全世界物理学家共享信息。

[2] 在1994年夏季至1995年间，Web浏览服务超越了文件传输服务成为Internet的主要用途，并且从那时开始一直在各项服务中位居首位。

Web 浏览器显示出指定的文档，并且允许用户选择突出显示的项目，这些项目可以由文本、图像、视频或音频组成。

23.15　超媒体显示的示例

可以用一个简单的例子来展示用户所见到的浏览器。当然，当浏览器在计算机屏幕上显示信息时，显示器看起来会明明亮亮多彩；这种效果是不能在书中简单再现的。商业的浏览器软件会在用于操作浏览器的按钮上显示图标，并且显示一个鲜艳的商标对浏览器的制造商加以宣传。在这里，我们不会展示商业浏览器的精切细节，而是给出一个与广泛使用的浏览器界面相类似的略图，并主要讨论浏览器提供的功能。

举个例子。思考浏览器会如何显示属于某个虚构的公司的 Web 文档。假设用户正在浏览万维网，并且遇到一个名为 Hypothetical Rocker Company 的公司。图 23-5 举例说明了浏览器会如何显示这个公司的 Web 网页[○]。

图 23-5　Web 浏览器所能显示的信息的虚构示例。其中任何突出显示的项目，包括文本和图像，都是可被选择的

23.16　浏览器的控制

理论上，浏览器与用户通过两种方式进行交互。浏览器与用户的第一种交互方式发生在浏览器获取了用户所请求的项目之后，浏览器自动显示相应的信息以供用户阅览。第二种交互方式发生在用户与浏览器通信时。例如，用户可以输入文档的地址，请求浏览器打印当前的文档，返回上一个文档，或者中止某个请求[○○]。

大多数浏览器都把显示窗口分成两个区域，每一个区域分别对应于一种交互方式。其中，主要区域包含用户正在浏览的文档，而周围的区域则包含了用户与浏览器进行交互的控制键。图 23-5 的例子说明了这一点。在该例中，浏览器将用户正在阅览的文档放在占据了大部分窗口的区域中。为了标示出界线，浏览器在该文档周围放置了边界，因此，用户可以立即分辨出窗口的哪一部分对应于该文档。

在该例中，浏览器使用窗口的上部与用户进行交互。在该区域中，浏览器安排了一些项目以使用户对浏览器的工作进行控制。例如，最上端的区域由一个水平工具栏组成，其中包含一系列术语：File，Edit，View，……，Help。每个术语都对应着一个下拉菜单（pull-down menu），每个菜单中都包含浏览器支持的一些命令。要想从列表中选择某个项目，必须移动鼠标，将光标移到该项目上，然后用鼠标单击此项。之后会出现一个选项菜单，允许用户使用鼠标作进一步的选择。

由于许多控制操作经常会用到，所以为这些操作设置了单独的按钮。这些按钮通常都位于

○　在现实中并不存在这个公司或 Web 网址。

○○　中止请求并不是常见的操作，这种情况通常在用户没有耐心等待浏览器传送较大文件的时候发生。

接近屏幕顶端的地方，头两个按钮是最常使用的：带着向左箭头的标为 Back 的按钮用于返回到上一个文档；而带着向右箭头的标为 Forward 的按钮用于前进到下一个文档。例如，假设某个用户按以下次序浏览了4个文档：A，B，C和D，如果屏幕上当前显示的是文档D，那么当用户选择 Back 按钮时会返回到文档C，再选择 Back 按钮时会返回到文档B。当浏览被打断时（例如被电话打断），按次序返回会尤其有用，因为这有助于用户重建浏览文档的前后关系。

23.17 外部引用

上例还说明了另一个概念：嵌入式菜单项可以引用公司外的信息。在图23-5所示的例子中，突出显示的短语 The Bakery 就指向 Hypothetical Rocker Company 所在街道对面的一家商店。也许，The Bakery 同意在 Hypothetical Rocker Company 的 Web 网页中提供对自己的引用，或者该引用是已经向 Hypothetical Rocker Company 付费而作的广告。

图23-6 对应于图23-5中项目 Types of rocking chairs 的一种示例显示

图23-5的例子显示了 Web 网页的一种可能的视觉效果，混合各种大小和风格的文本以及图像。表现形式的多样性以及一致性的缺乏会使信息看起来很杂乱，以至于难以阅读。

Web 网页并不需要做得太复杂。事实上，最好的项目应当简洁地组织和显示信息。例如，名为 Types of rocking chairs 的项目，Hypothetical Rocker Company 将其公司摇椅的信息存放在这个项目中。对每种摇椅来说，其信息可能是图片，或者是类似于本例所示的带有背景音乐的文档。当用户选择了项目 Types of rocking chairs 后，浏览器会获取并显示相应的信息。图23-6给出了一种将信息组织成简单列表的方法。

23.18 记录信息位置

假设你正在使用浏览器浏览万维网，并且发现一个网页中的信息让你特别感兴趣，你可能想把这个网页告诉你的朋友，或者记录下它的位置以便于以后浏览。由于在一系列超媒体文档中浏览的路径可能会很长很复杂，所以要想记住到达某个特定网页所进行的所有选择是很困难的。此外，由于 Web 文档并不是静态的，所以这次让你到达某个特定网页所作的一系列选择等到下次就未必有效。

为了使用户能够将文档的位置记录下来，万维网为每个网页都分配了一个惟一的标识符。该标识符由一串字符组成，并且能被记录在计算机文件中、写在纸上或者发送给另一个人。给定某个标识符，浏览器就能立即跳转到相应的网页。用户则不必在菜单中不断搜索和选择项目了。

用于指定某个特定 Web 网页的标识符被称为统一资源定位符（Uniform Resource Locator，URL）。当浏览器显示某个网页时，也会显示出该网页的 URL。URL 与电话号码类似，URL 是一个简短的字符串，用户可以记住它以避免搜索相应的网页。现在许多公司都将各自的 URL 放在广告中（例如，电视购物），或者附在产品标签上。

例如，图23-5中接近屏幕顶端的对话框中有一个 URL[⊖]：

⊖ 这个 URL 与该例中的公司一样，也是虚构的。

http://www.brocker.com/typev

URL 看起来像是一串字母和标点符号的无意义的组合，但恰恰是这样精确的语句表达出的含义使浏览器能够用其获取信息。

统一资源定位符由一个简短的字符串组成，用于标识特定的多媒体文档。给定一个有效的 URL，浏览器不用通过其他文档，就能直接跳转到相应的网页。

23.19　书签/收藏夹

为了便于用户直接打开经常访问的网页，浏览器中含有一种特殊的机制，一般称其为书签（bookmark）（有时也称为热单（hot list）或收藏夹（favorite））。该机制允许每个用户维护一个经常访问的 URL 的列表。例如，假设某个办公室职员需要经常访问另外三个公司（例如，三个供应商）的网页，该职员就可以为其中每个公司各建一个书签。

书签最主要的优点就是它既方便又快捷。在创建了一个书签列表后，用户就可以通过使用一个下拉菜单（如图 23-6 中名为 Bookmarks 的菜单）来选择该列表中的任何一个书签。当用户点击 Bookmarks 标签时，该用户的书签列表就会以菜单形式出现。如果用户点击了该列表中的某个项目，那么浏览器就会获取并显示该项目所对应的 Web 网页。

书签具有永久性。当用户退出浏览器或关闭计算机时，书签并不会消失。永久性是通过将书签列表存储在计算机磁盘上的文件中来做到的。一旦书签列表有了改变，浏览器就将其保存在磁盘上；无论何时浏览器开始运行，都从磁盘上读出该列表。因此，书签一旦被保存，以后就能一直使用。

维护书签列表与使用它一样方便。通过使用鼠标进行一系列选择可以更新书签列表：用户并不需要手工键入 URL。也就是说，当用户发现了他感兴趣的网页时，可以选择相应的下拉菜单，并选择项目收藏本页（Bookmark This Page）或添加新书签（Add New Bookmarks）。浏览器就会将当前的 URL 加入书签列表，便于以后快速地访问该网页。与此类似，用户可以选择编辑书签（Edit Bookmarks）或管理书签（Manage Bookmarks）来删除已有的书签或改变书签在下拉菜单中的显示方式。

23.20　万维网如何工作

与早期的浏览服务一样，万维网也使用客户机/服务器的交互方式。Web 浏览器作为客户机，通过 Internet 与远程 Web 服务器进行联系以获取用户请求的网页。远程系统上的服务器会返回该网页的副本以及其他的附加信息。

Web 服务器返回的附加信息向浏览器传递了两个重要信息。第一，它描述了信息应当如何显示。第二，它给出了网页上每个可选择项目的 URL。

当浏览器从远程服务器获取了某个网页后，就将它显示出来，并且等待用户选择其中的一个突出显示的项目。一旦用户作出了选择，浏览器就会查询与该网页一起到来的附加信息，并找到用户所选项目的 URL，然后，浏览器会通过 Internet 获取新选择的网页。

23.21　URL 告诉浏览器所要联系的计算机

浏览器是如何知道哪台远程计算机中含有某个特定网页的呢？URL 会告诉它。每个 URL 所包含的信息中都含有相应远程计算机的域名。URL 与电话号码相似：URL 是一个简短的字符串，用来指定一个特定的联系点。当给定某个 URL 时，浏览器从中提取出相应的计算机域名，然后通过 Internet 与该计算机上的服务器联系。

23.22 URL告诉浏览器所要联系的服务器

除了远程计算机的域名外，URL还指定了特定的服务器以及能从该服务器上获取的一个网页。在与相应的远程计算机上的服务器连接上后，浏览器向该服务器要求得到指定网页的副本。总之：

> 每个URL通过给定一台远程计算机的域名，该计算机上的一个服务器和该服务器上的一个特定网页来惟一确定一个网页。

图23-7举例说明了URL如何编码相关的信息。URL的第一部分指定了访问协议（access protocol），浏览器使用它来与远程服务器联系。在冒号和双斜线之后是服务器所在的计算机的域名。最后，使用单斜线将该计算机域名和代表某个特定项目的后缀分隔开。

虽然图23-7中的示例URL说明了一个典型的URL所含有的各个部分及每部分的含义，但是并非所有的URL都具有相同的格式。首先，许多浏览器都允许用户省略协议前缀；如果没有给定前缀，浏览器会在相应的URL前加上"http://"。其次，许多URL也会省略掉后缀。每个Web服务器都含有一个默认项目；当没有指定任何项目时，服务器则会返回该默认项目，服务器的默认项目通常是一个网页，该网页提供了对所有可得信息的总览和摘要，以及指向其他网页的链接。因此，有些浏览器允许URL中只包含服务器的域名。

图23-7 URL各部分的含义，并非所有的URL都需要有所有这些部分

23.23 URL中"www"的使用

由于Web服务器可以在任何接入Internet的计算机上运行，所以URL中的域名并不必以www开头。那么为什么大多数URL中的域名都以www开头呢？大多数组织为运行着该组织Web服务器程序的计算机取名为www是为了使URL简单一致，易于记忆。

将运行着Web服务器的计算机命名为www是非常容易的。这是由于域名系统允许网站为一台计算机分配多个名字。例如，当Purdue大学的计算机科学系安装Web服务器时，在已有的计算机中挑选一台来运行该服务器程序，该计算机的域名为lucan.cs.purdue.edu，IP地址为128.10.19.20。为了便于人们找到运行着服务器的计算机，该系在域名系统中为该计算机注册了另一个域名www.cs.purdue.edu。因此如果用户输入了URL：

http://www.cs.purdue.edu

浏览器会联系与输入以下URL后所联系的相同的服务器：

http://lucan.cs.purdue.edu

23.24 浏览器提供对多种服务的访问

除了提供对多媒体文档的访问外，万维网还从另一个重要方面对早期的浏览服务进行了扩展：它提供对多种服务的访问。例如，Web浏览器可以从本地磁盘文件中，或从FTP文件传输服务（将在第29章中介绍）中访问信息。此外，Web文档中的可选择项目可以引用其他服务的信息，就像引用其他Web文档一样。

理解Web通用性的关键是要认识到URL机制会指定服务类型以及运行着该服务的计算机的域名。以"http"开头的URL表示用户使用Web服务来检索标准的Web文档。以"ftp"开头的URL表示一个可以通过FTP文件传输服务访问的文件。

人们经常使用与Web服务相同的技术来编码其他服务。例如，商务名片中通常会含有以下内容：

Telephone（电话）：　315-555-7895
Cell（手机）：　315-555-9537
Fax（传真）：　315-555-7802

其中，起始的字符串指定了该号码是用于拨打语音电话的，还是用于发送传真的。也就是说，像"Fax:"这样的前缀是用来指定所使用的访问机制的。

23.25　浏览器程序的内部

由于URL可以指定所使用的服务类型和待检索项目的位置，所以浏览器必须为处理多种服务做好准备。也就是说，浏览器程序必须包含访问各种服务所需的所有软件。从理论上看，可以将浏览器看作是一个内置若干个客户机程序的庞大程序。浏览器的接口部分通过接收输入和控制浏览器支持的操作来与用户进行交互。当用户输入一个URL或从文档中选中一个链接时，浏览器就提取出该URL中用于指定服务类型的前缀，并使用该前缀来选择相应的客户机程序。然后，接口程序调用该客户机程序来获取指定的项目，并向用户显示出结果。图23-8举例说明了这一概念。

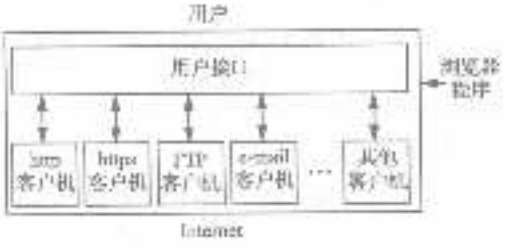


图23-8　浏览器的基础结构，用户与一个单独的，统一的接口进行交互，该接口使用URL中的信息来选择一个内置的客户机程序，并用它来访问Internet上的信息

例如，浏览器可以发送邮件，并允许用户查看新到来的电子邮件。浏览器还可以使用FTP文件传输服务来将文件下载到用户的计算机上。从用户的角度来看，相应机制的选择是自动的：用户浏览万维网，通过作出选择来请求信息；浏览器利用URL来决定如何继续工作。

> 浏览器之所以新颖和功能强大，是因为它将对多种Internet服务的访问集成为一个单独的、无缝的浏览系统。浏览器使用URL中的信息来自动地从各种服务（如电子邮件服务、文件传输服务）中选择相应的访问机制。

23.26　小结

Internet浏览服务从两个重要的方面增强了对信息的访问能力。首先，这种服务使用交互的方式，使用户能够浏览远程计算机上的文档，而无须进行乏味的文件传输。更重要的是，浏览服务将可选择的项目，如文本、音频、视频、图像等，集成到文档中。结果就形成了超媒体系统，该系统中的多媒体文档可以包含对其他多媒体文档的引用。为了显示多媒体文档的输出结果，计算机中必须含有显示图像和再现声音的硬件。

万维网已经成为Internet上最为流行的信息服务。用户可以将Web看作是存储在遍及Internet的计算机上的大量超媒体文档。一个给定的Web文档中可以包含指向存储在其他计算机上的Web文档的嵌入的菜单项。用户能够方便地由一台计算机上的一个文档链接到另一台计算机上的文档上，就像在同一台计算机上的文档之间进行链接一样。

浏览器是用于访问万维网的计算机程序。浏览器提供点击式的界面，允许用户使用鼠标来访问信息。在浏览器显示了某个特定的多媒体文档后，用户可以通过点击嵌入式菜单项来请求浏览器获取和显示新的文档。因此，从用户的角度来看，信息所处的位置并不重要：万维网看起

来是文档间的无缝连接，其中每个文档都能被直接访问。

为了能够记录某个特定信息的位置，每个 Web 文档都被分配了一个惟一的名字，称为统一资源定位符（Uniform Resource Locator，URL），其中含有浏览器获取文档所需要的信息。URL 可以由用户手工输入，也可以从嵌入式链接中获得。给定一个 URL，浏览器就能够确定访问相应文档的服务类型、该文档所在的计算机，以及该计算机上特定项目的位置。由于浏览器中包含了用于各种访问机制的软件，所以它能够访问菜单系统中的项目，提供下载的文件，或者 Web 服务器上的超媒体文档。因此，浏览器提供了一种统一、无缝的方式来访问 Internet 上的大多数信息。

23.27 对超媒体浏览的评论

大多数人都觉得万维网异彩纷呈。在短短几分钟内，用户就可以学会如何进行选择，以及如何从一个网页跳转到另一个网页。不用多久，初学者就会发现他们已经浏览了有关许多主题的文档，而这些文档位于遍及全世界的 Internet 网站中的许多计算机上。例如，某个中西部大学的一位教授会看到他将要游览的夏威夷的彩色图片，而他的同事可以在欧洲的某个网站上浏览一部外国电影的预告片。

Web 文档之间互相连接的复杂性也在不断地发展变化着。由于万维网正以难以置信的速度增长着，所以 Web 中所包含的信息也在不断地变化，这会诱使用户不断地去浏览。结果是，Web 浏览可能会使人上瘾，并让人消磨掉很多时光。

第 24 章　万维网文档（HTML）

24.1　引言

第 23 章介绍了万维网，并讨论了用户眼中的信息浏览服务。本章将介绍 Web 页面的内部表示方法。首先介绍创建 Web 页面必须使用的语言，然后解释如何创建一个包含文本、图像的多媒体文档，以及一个 Web 页面中的链接如何指向另一个网页。

为什么要学习 Web 页面的内部表示法呢？毕竟，浏览器已经向用户隐藏了所有内部细节。这有两个原因，首先，学习一些基本概念有助于理解超媒体的思想，并消除 Web 页面的些许神秘感。其次，学习 Web 页面的内部语言可以使用户了解表达一个 Web 页面需要哪些细节。

24.2　显示硬件的改变

计算机的显示硬件差异很大，一些计算机的监视器的分辨率较低，只能显示 256 色；另一些计算机可以显示百万色。一些显示系统只能显示几种字体，而另一些系统则能以任意大小和各种字体显示文本。更重要的是，Internet 的流行意味着全球的用户会使用各种类型和样式的计算机来访问和显示 Web 页面。

这样看来，为万维网设计多媒体信息网页的人需要了解许多类型计算机的详细信息。毕竟，Web 页面的设计者必须指定如何表示 Web 页面及其内容，也就是说，设计者必须指定每条显示在网页上的消息的文本，是否要突出显示某条消息（例如，以比周围文本更大的字体来显示），要加入哪些图片，以及网页中的项目布局。

24.3　浏览器解释并显示 Web 文档

有趣的是，创建 Web 页面的人既不需要了解计算机显示硬件的详细信息，也不需要为每种类型的计算机显示屏幕准备不同的网页副本。相反，网页是以一种通用的方式创建的，并且能够适用于任何显示硬件。因此，设计者只需要为每个网页准备一个单独的副本，而不需要在网页说明中指定如何在各种不同类型的计算机硬件上显示该网页。设计者使用一种计算机语言来详细说明网页的内容。这种语言包含了用于显示的标准，而不给定细节。例如，设计者可以指定组成一段话的一组句子，但却不能指定每行长度或者是否缩排段的开头这样的细节。

网页中的标准是如何在一台特定计算机的显示硬件上产生输出的呢？Web 浏览器中的软件进行了必要的翻译。当浏览器从 Web 上获得一个网页时，它收到一份与设计者所写的网页说明完全一样的副本，并无设计者的标准翻译成适于本地计算机显示硬件的具体细节。

例如，浏览器知道用于显示文本的窗口的尺寸，并利用这个尺寸来决定每行安排多少个单词。同样地，如果某台计算机能显示彩色，并且 Web 页面指定了用红色显示某单词，那么该计算机上的浏览器就会指示硬件将该单词显示为红色。如果某台计算机不能显示彩色，那么该计算机上的浏览器将以黑白方式显示网页，并用不同灰度的灰色来代替彩色。从本质上说，浏览器会力图在已知硬件上创造最好的显示效果。

24.4　Web 所用做法的结果

允许浏览器来选择显示细节会产生一个有趣的结果：当使用两个不同的浏览器或者在两台硬件不同的计算机上浏览同一个 Web 页面时，显示效果有可能不同。除了彩色与黑白的差异外，有些计算机的显示屏幕会比其他计算机的更宽一些，因此，在一行中能够显示的文本长度或图片尺寸会因计算机的不同而有所不同。为了弥补这些差异，浏览器必须能够将一行长的文本分在两个物理行中显示，或者只显示一页的一部分并允许用户移动可视部分来浏览其余的部分。总之：

> Web 页面是用一种计算机语言编写的，该语言给出了所要显示内容的基准；当在某特定的计算机显示器上显示某 Web 页面时，浏览器将该网页的说明翻译成适用于该显示硬件的命令。因此，当在两台不同的计算机上或使用不同的浏览器显示同一 Web 页面时，会出现微小的差别。

24.5　HTML：用于 Web 文档的语言

虽然 Web 页面使用的语言是高级语言，但它并不是诸如英语之类的自然语言。Web 页面是由超文本标记语言（HyperText Markup Language，HTML）编写而成的。

与计算机使用的其他语言一样，HTML 有它自己的语法规则，并且以特有的方式使用普通的标点符号。遗憾的是，HTML 是为了便于计算机处理而设计的，因此，这些语法规则使得人来编写 HTML 文档变得很困难。事实上，要阅读或理解 HTML 也是困难的[⊖]。

习惯于使用文字处理程序的人会惊讶地发现，Web 页面的说明在设计者创建时和被浏览器翻译显示后并不相同。事实上，编写出 Web 页面的 HTML 语言与浏览器显示的结果页面有着明显差不同。例如，HTML 文档中每行中单词的排列与浏览器显示该文档时每行中单词的排列没有任何关系。图 24-1 中，通过对比设计者用 HTML 创建的 Web 页面的一部分和浏览器显示该文档时的显示结果来说明这个思想。

a)　　b)

图 24-1　a) 用 HTML 创建的 Web 页面的一部分；b) 浏览器显示此网页的结果。HTML 中的空格与网页显示时的空格毫无关系

在这个例子中，作者把组成一个句子的几个单词分隔在几行上，并在单词之间加入任意数量的附加空格。在浏览器显示此文档时，所单词都显示在一行中，并且单词之间已经没有不必要的空格了。

允许 HTML 包含任意多个空格和附加行的概念就是所谓的自由格式输入（free format input）。使用自由格式输入的目的很明确：由于 HTML 并不明确指定显示器输出结果，所以浏览器可以自由选择适合给定计算机的输出格式。

24.6　Web 页面中的指令能够控制输出

尽管 HTML 使用自由格式输入，但是 Web 页面中可以加入一些指令来指导浏览器如何解释和显示该网页。例如，设计者可以在 Web 页面中加入一些指令，来告诉浏览器何时开始输出新

⊖ 要查看某网页的 HTML，先用浏览器显示该网页，然后点击 View 菜单下的 Page Source 选项。

的一行，何时把一行文本居中显示，以及将项目以列表形式显示还是以段落形式显示。

为了把 HTML 指令和要显示的文本分隔开来，每条指令都是由被小于号（<）和大于号（>）括起来的字符串组成的。这样的指令被称为标记（tag）。例如，HTML 的标记
指示浏览器开始新的一行[1]。使用小于号和大于号来区分标记和普通文本的效果很好，这是由于正常的英语语法不会在单词周围放置这些符号。

24.7　Web 页面分为两个主要部分

每个 HTML 文档都分为两个主要部分：头部（heading）和正文（body）。头部包含用于标识网页的信息，正文则包含了将要显示的实际信息。例如，头部包含了网页的标题。浏览器可以选择是否以及如何显示头部项目；大多数浏览器会在屏幕的一块单独区域中显示头部信息，而将主要的浏览器窗口预留给文档正文中的项目。

设计者使用标记来告诉浏览器网页的哪一部分对应于头部，哪一部分对应于正文。与 HTML 中的许多标记一样，用于括住头部和正文的标记都是成对出现的，其中一个用来开始一部分，另一个用来结束这部分。除了结束标记的名字前面有一条斜线外，成对出现的标记的名字是相同的。例如，标记对<HEAD>和</HEAD>括住了头部。类似地，标记对<BODY>和</BODY>括住了正文。另一个标记对<TITLE>和</TITLE>出现在头部内，它们括住了文档标题。

图 24.2 举例说明了一个典型的 Web 页面中标记对的排列。当然，浏览器不会把标记显示给用户看。标记仅仅控制浏览器显示输出的方式。

```
<HTML>
<HEAD>
    <TITLE>
      ... Formal Document Title Goes Here
    </TITLE>
    ... additional heading information, if needed
</HEAD>
<BODY>
    ... actual information to be displayed on the page
</BODY>
```

图 24.2　将 Web 页面分为头部和正文的基本标记[2]。浏览器把头部和正文分开显示

24.8　缩排使 HTML 易读

如前所述，HTML 使用自由格式输入，允许插入任意数量的空格。但是，图 24.2 例子的格式却很规整。每个标记都放置在不同的行上，并且头部和正文内部的信息都被缩排了。此外，像<HEAD>和</HEAD>这样的标记对的缩进量是完全相同的。

为什么设计者要如此费神地把 HTML 标记放置在不同的行上并且缩排每个部分呢？当然不是为了方便浏览器：当浏览器读取网页时，会忽略所有多余的空格。答案是：精心地排列 Web 页面的说明是为了让人们更容易理解它们。除了编辑 HTML 文档外，可能还需要人工修改出错的文档或更新信息变化了的文档。将 HTML 标记放置在不同的行上和缩排文本可以使人们更容易把握文档结构和定位个别项目。

---

[1] 使用 BR 的原因是印刷行业使用技术术语空白中断（line break）来表示新的一行的开始。

[2] 本节这个例子中的标记都是大写的，但是 HTML 并不区分大小写。因此，<head>与<HEAD>是等价的。

人们很少查看 Web 页面的内部格式。事实上，人们可以使用许多自动化工具来创建 Web 文档，而不用键入或查看内部 HTML。但是，专业程序员有时需要人工修改 HTML 来改进自动化程序生成的代码。总之：

> 尽管浏览器在显示 HTML 文档时忽略了缩排，但是将标记放置在不同的行上和缩排每一项可以使人们更容易地阅读 HTML 文档。

24.9　在 Web 页面的正文中加入文本

如果 Web 页面中包含的文本被组织成几个段落，浏览器会在显示该文本时在各段之间插入一个空行。但是，浏览器并不希望在网页的 HTML 版本中用空行分隔段落。设计者需要使用段落标记来告诉浏览器一段的结束和另一段的开始。段落标记由 3 个字符组成：<P>，选择 P（paragraph）是为了便于记忆。

图 24-3 举例说明了一个 HTML 文档的一部分，包含 3 个段落，并给出了浏览器显示该文档时的输出结果。

```
This is an example of text on a Web page arranged in paragraphs. The first paragraph is quite short; the
paragraph contains only two sentences.
<P>
    The second paragraph contains two sentences. A sentence can span multiple input lines and can con
    tain additional    spaces; a browser ignores each spacing when displaying the paragraph.
<P>
    This is the final paragraph. In HTML, a tag separates each pair of paragraphs. On the screen, how-
    ever, vertical blank space separates paragraphs as in a textbook.
```

a)

```
This is an example of text on a Web page arranged in paragraphs. The first paragraph is quite short; the paragraph con-
tains only two sentences.

The second paragraph contains two sentences. A sentence can span multiple input lines and can contain additional spaces; a
browser ignores each spacing when displaying the paragraph.

This is the final paragraph. In HTML, a tag separates each pair of paragraphs. On the screen, however, vertical blank
space separates paragraphs as in a textbook.
```

b)

图 24-3　a）一个 HTML 文档的一部分，包含 3 个文本段落；b）浏览器显示该文档的结果。浏览器通过 HTML 文档中的标记来决定如何把句子组织成段落

24.10　缩排使段落更易被发现

图 24-3a 的示例 HTML 已经经过格式化，以使人们更容易阅读。设计者仔细地将每个段落标记放置在单独的行上并且缩排文本。前面讲到，HTML 的缩排格式是为了方便人们阅读而设计的，并不影响浏览器显示网页。例如图 24-4 举例说明了图 24-3a 中的 HTML 文档的另外两种可能的格式。虽然浏览器显示这二种版本的 HTML 文档的结果是完全相同的，但是图 24-3a 中的版本更易于人们阅读和修改，因为 HTML 文档中的空格正好与浏览器显示该文档时的空格相对应。

```
This is an example of text on a Web page arranged in paragraphs. The first paragraph is quite short; the
paragraph contains only two sentences. <P> The second paragraph contains two sentences. A sentence
can span multiple input lines and can contain additional spaces; a browser ignores each spacing when
displaying the paragraph. <P> This is the final paragraph. In HTML, a tag separates each pair of para-
graphs. On the screen, however, vertical blank space separates paragraphs as in a textbook.
```

a)

图 24-4　图 24-3a 中的 HTML 文档的另外两种版本。浏览器显示这二种版本的结果如图 24-3b 所示

```
This is an example of text on a Web page arranged in paragraphs.

The first paragraph is quite short; the paragraph contains only two sentences. <P> The second
paragraph contains two sentences.

A sentence can span multiple input lines and can contain additional spaces; a browser ignores such
spacing spacing when displaying the paragraph. <P> This is the final paragraph. In HTML, a tag separ-
ates each pair of paragraphs. On the screen, however, vertical blank space separates paragraphs as in
a textbook.
```

b)

图 24-6　(续)

24.11　一个 Web 页面能够链接到另一个 Web 页面

第 23 章介绍了可以在 Web 页面中加入可选择链接，允许用户通过使用鼠标点击一个项目来从一个 Web 页面转移到另一个 Web 页面。Web 页面的设计者如何指定一个项目与一个可选择链接相对应呢？HTML 使用一个标记对：在创建 Web 页面时，将这样的一个标记对分别放置在每组组成可选择链接的项目前后，这样的标记对可以放在一个单词、两个单词、一个短语或者一整个段落的前后。当浏览器发现这样的标记对时，会把标记对之间的项目标记为可选择链接。通常，浏览器在这些项目下面加上下划线并以彩色显示它们。

在 HTML 术语中，网页中对应于链接的项目称为被锚定（anchored）了。自然地，用字符 A 在一个标记对中标记链接。一个链接以标记<A>开始，以标记</A>结束。

为了指定某个特定链接所指向的网页，开始标记中要加入关键字 HREF，后面跟着一个等号和一个用双引号括起来的 URL 地址。例如，图 24-5 举例说明了一个 HTML 文档的一部分和其对应的浏览器显示输出结果。在这个例子中，短语 William Shakespeare 被锚定到 URL 地址：

http://the-tech.mit.edu/Shakespeare

当浏览器显示锚定文本时，会改变锚定项目的颜色并在其下面加上下划线，来告诉用户该文本与可选择链接相对应。当用户点击了某个可选择项目后，浏览器会载入由其 URL 地址指定的新网页。

```
On the Web, you can find some of the classic works of English literature. For example,
the works of
<A HREF="http://the-tech.mit.edu/Shakespeare">
                William Shakespeare
</A>
are available.
```

a)

On the Web, you can find some of the classic works of English literature.
For example, the works of William Shakespeare are available.

b)

图 24-5　a) 一个 Web 页面的部分 HTML 文档，包含一个指向另一个网页的链接；b) 浏览器的显示结果，大多数的浏览器以不同的颜色或以灰色来显示链接

24.12　HTML 支持已编号的和未编号的列表

HTML 包含了相应的机制来格式化项目。例如，图 24-6 举例说明了用于创建有序列表（ordered list）的 HTML 文档，有序列表有时也称为已编号的列表（numbered list），标记对<OL>和</OL>分别位于整个列表的前后，而标记<LI>则位于每个表项之前。

```
<OL>
    <LI> breakfast
    <LI> lunch
    <LI> dinner
</OL>
```

a)

1. breakfast
2. lunch
3. dinner

b)

图 24-6　a) 用于创建有序列表的示例 HTML 文档；b) 浏览器的显示结果

对应于有序列表，HTML 也支持无序列表（unordered list），或称为带项目符号的列表（bulleted list）。如图 24-7 所示，无序列表也在每个表项前面使用标记<LI>，但是在列表前后使用标记对<UL>和</UL>。

```
<OL>
    <LI> cars
    <LI> trucks
    <LI> buses
</OL>
```

a)

* cars
* trucks
* buses

b)

图 24-7　a) 用于创建无序列表的示例 HTML 文档；b) 浏览器的显示结果

24.13　Web 页面中的图像是数字的

到目前为止，我们已经介绍了在 Web 页面中加入文本的方法。然而，HTML 最让人感兴趣的地方之一是它不仅能包含文本，还能包含图片或图形。在 Web 页面中加入某图像之前，该图像必须经过数字化（digitized），也就是说，该图像必须被转换为一个数字序列，并存储在磁盘上的一个数据文件内。

数字图像中的每一个数字都代表着一小片相应图片的信息。这些片极其微小，典型的数字图像中单位英寸的面积上含有几百个这样的小片。当浏览器在计算机屏幕上显示数字图像时，就使用数据文件中的数字序列来确定屏幕上像点的颜色和亮度。

计算机屏幕上的数字图像与电视图像有着相同的工作原理：人眼分辨微小事物的能力有限，当每英寸屏幕上有几百个像点时，人眼是不能分辨出单个像点的。人类通过大范围的颜色和明暗变化来分辨显示的图像。人在看到计算机上的数字图像或电视屏幕上的图像时，只能分辨出图像，而不能分辨出每个像点。

数字图像是怎样生成的呢？通常有两种方法。第一种方法是使用数码相机（digital camera），第二种方法是使用数字扫描仪（digital scanner）。在数码相机中，镜头将光聚焦在一个电子设备上，该设备将图像转换成数字形式，数字图像可以被发送到计算机上，并存储在磁盘上的文件内。

第二种方法使用数字扫描仪，数字扫描仪是连接在计算机上的一种设备。用户把图片或文档放入扫描仪中，并开始扫描，扫描仪机械图片扫描，生成数字版本，并把扫描结果发送到计算机上。与数码相机一样，扫描结果存储在计算机磁盘上的文件内。图 24-8 说明了生成数字图像的两种方法。

与普通的照片能够存储为底片、彩色幻灯片或彩色印刷品一样，数字图像也能以几种不同的格式进行存储。从本质上说，所有格式都表达着相同的信息，但是存储格式的细节各有不同。此外，需要安装一定的计算机软件来解析某些特殊的图像格式，使用为另一种格式设计的程序来解析当前格式，就像使用幻灯片放映机来放映照片底片一样。

最流行的数字图像存储格式之一是交互式图形格式（Graphics Interchange Format，GIF）[1]。另一种格式是联合图像专家组（Joint Photographic Experts Group，JPEG）。尽管现在有计算机软件进行数字图像格式之间的转换，但是一些扫描仪却只能生成一种特定格式的数字图像。

图 24-8　a）数字扫描仪示例；b）数码相机示例，两者都能生成存储在计算机磁盘上的图像文件

24.14　HTML 支持在 Web 页面中加入图像

当一个 GIF 图像文件存储在某台计算机的磁盘上后，该计算机上的 Web 页面就能通过引用该文件来加入相应图像。在 HTML 中，图像的引用使用 IMG 标记，并要加入关键字 SRC 和磁盘上的文件名。例如，假设某人使用数字扫描仪生成了一个 GIF 图像，并且该数字图像被存放在计算机磁盘上的文件"xxx.gif"中。在 Web 页面中加入该图像的标记为：

```
<IMG SRC="xxx.gif">
```

注意，所有对图像的引用都要使用关键字 IMG 和 SRC；只有每个图像的文件名是不同的。

文件的名字取决于计算机所使用的操作系统。在一些系统中，文件名以斜线开始，并且包含多个以斜线分隔开的单词。在另外一些系统中，文件名以单个字符开始，后面跟着一个冒号。例如，如果一个数字图像文件名为 r:stickman.gif，那么该文件所在计算机上的 Web 页面可以使用下面的标记来加入该图像：

```
<IMG SRC="r:stickman.gif">
```

24.15　文本可以出现在图像周围

Web 页面上的图像可以单独出现，也可以出现在文本周围。事实上，当浏览器显示 Web 页面时，它把图像当作是出现在一行文本中的一个巨大的"单词"。例如，图 24-9 举例说明了浏览器如何显示一行文本中加入的图像。

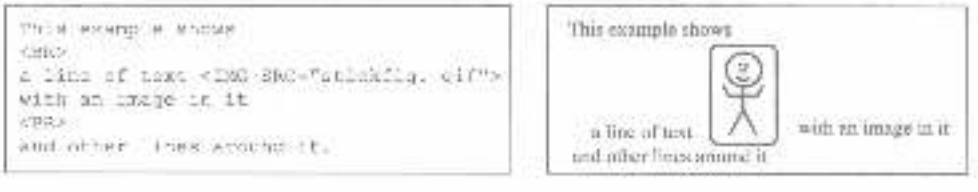


图 24-9　a）一个 Web 页面的一部分，既包含文本又包含图像；b）浏览器显示该网页的结果，浏览器把图像当作是巨大的单词

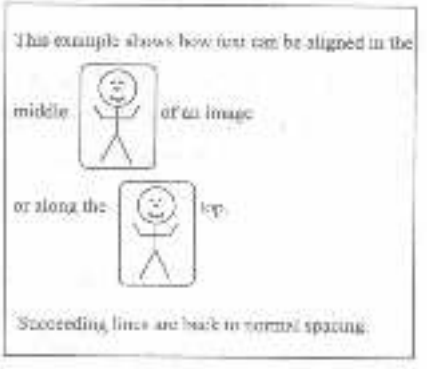
HTML 提供了如何控制图像与周围文本对齐的方法。对齐信息在 IMG 标记中，如果不指定对齐信息，那么图像的底部与文本在同一行，如图 24-9 所示。如果 IMG 标记指定 ALIGN=TOP，那么浏览器将把图像的顶部与文本放在同一行；如果 ALIGN=CENTER，那么浏览器将把图像居中放置。图 24-10 给出了 IMG 标记中的对齐说明和显示结果。

---

[1] 大多数专业人员将 GIF 读为类似于"gift"的发音。

```
This example shows how text can be aligned in the
<BR>
middle<IMG SRC="stickfig.gif" ALIGN=CENTER> of an image
<BR>
or along the<IMG SRC="stickfig.gif" ALIGN=TOP> top.
<BR>
Succeeding lines are back to normal spacing.
```

a)



b)

图 24-1b　a）一个包含图像和文本的 Web 页面的一部分；b）浏览器如何解释对齐说明的示例

24.16　图像能够链接到另一个 Web 页面

我们已经知道，在文本前后放置标记对<A>和</A>可以使该文本成为指向另一个 Web 页面的可选择链接。HTML 使用相同的标记对来使图像成为可选择链接。例如，假设一个数字图像被存放于文件 yyy.jpeg 中，下面的 HTML 会在 Web 页面中加入该图像，并使该图像成为一个可选择链接：

```
<A HREF="http://www.xxxhaz.com/index.html">
    <IMG SRC="yyy.jpeg">
</A>
```

如果用户点击该图像，那么浏览器会根据相应链接转移到 URL 地址：

http://www.krocker.com/index.html

24.17　有些浏览器能放大或缩小图像

数字图像可以被放大或缩小。更重要的是，改变数字图像的尺寸并不需要大量的处理工作，因此构建一个能在数字图像显示的同时改变其尺寸的浏览器是完全可能的。为了使改变图像尺寸更加便捷，HTML 支持在 IMG 标记中加入更多的信息来指定期望的图像高度和宽度。

遗憾的是，不是所有的浏览器都能改变图像的尺寸。有些浏览器会忽略图像高度和宽度的说明，而以图像创建时的原始尺寸来显示图像。因此，为了保证 Web 页面中的图像在所有浏览器中都能以同样的尺寸显示，Web 页面通常都加入尺寸已经是期望值的图像，而不依靠浏览器来改变图像的尺寸。

24.18　背景控制

除了在 Web 页面中加入文本和图像外，HTML 还允许设计者指定 Web 页面的背景。这有两种方法：设计者指定背景的颜色，或者指定一个图像作为背景。如果为背景指定了颜色或图像，那么浏览器首先用该颜色或图像填充整个网页，然后再开始添加文本和图像。

背景图像可能会削弱显示效果，使阅读信息变得困难。但是，人们通常会选择能够增强显示效果的背景图像。因此，背景图像通常会使用柔和的颜色，而不使用有复杂细节和鲜艳颜色的图片。例如，背景图像通常会使用纹理样式，网页显示在这样的背景上就像文档打印在有纹理的纸上一样。

前面提到，每个浏览器都能选择显示网页的窗口的大小和形状。如果背景图像不能覆盖整个窗口会出现什么情况呢？浏览器会简单地对背景图像进行复制，与铺瓦片类似，这种技术被称为平铺窗口（tile the window）。因此，为了使整个背景都出现纹理，只需要创建一个包含所需纹理的小块矩形图像，再让浏览器使用其副本平铺整个屏幕即可。

24.19　HTML 的其他特色

HTML 还支持许多本章没有介绍的功能。例如，HTML 允许在 Web 页面中加入二维的信息“表格”。当浏览器显示表格时，会对齐相应的行和列。HTML 还提供了一种方法把多个可选择链接与一个图像联系起来，使每个可选择链接能分别对应于图像中的一个项目。例如，某个百货公司的 Web 页面中包含该公司的一个图片，图片中用不同的标志代表各个部门。如果用户将鼠标移动到显示某特定部门的区域并选中它时，浏览器则会显示该部门的相关信息。还有，Web 文档中能加入视频剪辑或在显示 Web 页面时播放音频剪辑。

当然，使用高级功能所需的 HTML 要比本章中的示例更加复杂。但是高级功能还是使用相同的模式：在输入中嵌入标记来指定如何显示相应项目。

24.20　HTML 的重要性

在 HTML 出现之前，在计算机屏幕上同时显示文本和图像需要很复杂的计算机程序。因此，只有具有大量计算机编程经验的人才能使图像显示在屏幕上。

HTML 看起来需要高级的技术技巧，但其实并不需要。大多数设计 Web 页面的人员只接受过平面艺术和设计的训练，而无计算机科学的专业背景。其中几乎没有人是计算机程序员。

24.21　使用 GUI 工具创建 Web 页面

虽然 HTML 是 Web 语言，但是已经有工具能够隐藏其内部细节，使人们不用学习 HTML 就能编辑 Web 页面。这样的工具称为 Web 设计工具（Web authoring tool），它们都有类似于文字处理程序的图形用户界面（Graphical User Interface，GUI）。当用户创建网页时，Web 设计工具会给出该网页在浏览器中的显示结果。例如，如果用户选择使用某种背景颜色，那么 Web 设计工具就会将背景改变为该指定颜色。当用户输入文本时，Web 设计工具会进行分行，并且将这些行显示在与浏览器显示它们时几乎一样的位置。最后，当用户加入某图像时，Web 设计工具会以与浏览器几乎一样的方式显示该图像。Web 设计工具允许用户通过移动、删除或修改项目来编辑网页。当网页显示在屏幕上后，用户可以将其保存成文件。

Web 设计工具与文字处理程序之类的应用程序最大的不同之处在于：网页的保存版本是由 HTML 写成的。因此，Web 设计工具的输出可以直接被浏览器读取。当然，由于 HTML 只给出

了显示的基准，所以用户创建网页时的显示效果与浏览器的显示结果会有微小的差别。要点是：

尽管HTML是主要的底层语言，但是已经有Web设计工具允许用户不学习HTML也能编辑Web页面。

24.22 小结

编写Web页面的计算机语言称为超文本标记语言（HyperText Markup Language，HTML）。由于浏览器在显示Web页面时完全隐藏了网页说明，所以大多数用户从来没有接触过HTML。当用户在浏览器中输入URL地址后，浏览器与相应的服务器进行连接，获得该网页的一个副本，解释HTML，并显示相应的结果。

HTML之所以重要，有两个原因。首先，HTML非常通用，可以在各种各样的浏览器和计算机上使用。特别地，HTML并不指定所有的显示细节，而是允许浏览器自由选择。其次，HTML使没有计算机编程背景的人也能创建Web页面。万维网的流行在很大程度上应当归功于HTML，并且Web设计工具使得不论个人还是公司创建Web页面都变得非常容易。

第25章　高级网页技术（表单、框架、插件、Java、JavaScript脚本和Flash动画）

25.1　引言

第24章我们介绍了万维网（World Wide Web）和用于创建普通网页的语言。本章将进一步介绍一些高级网页技术，应用这些技术能够给用户带来更好的交互体验，如动画、音频和用户间的直接对话等。本章首先介绍浏览器如何加载普通网页，然后进一步解释这些高级应用的实现过程。

25.2　传统Web采用静态页面

很多网页都是利用第24章所介绍的技术制作而成，因其内容不可改变，所以又称为静态（static）或被动（passive）网页，这就好像书中的一页内容，不会改变。

静态网页并不区分来自不同用户的请求，如果两个人在完全相同的计算机系统上，用同样的浏览器来访问同一个静态网页，他们将看到同样的内容：文字和图片一样，链接也都指向相同的页面。

静态网页的另一个特性是：用户在不同时间看到的网页内容也一样。例如，一个用户今天访问了某个静态网页，明天他再次访问该网页时，还会看到同样的内容。当然，如果网页的作者对网页进行了修改（例如更正错别字），用户再次访问时就可以看到修改后的新内容。但在作者对静态网页进行下一次修改之前，它的内容都不会再改变。

总结起来：

> 绝大多数网页都是静态的，这就意味着所有用户访问该网页时看到的内容都是相同的。

25.3　服务器如何存储静态页面

如果企业或个人想要在万维网上发布信息，必须具备三个条件：

- 一台带有硬盘的计算机
- 计算机与Internet之间的连接
- Web服务器软件

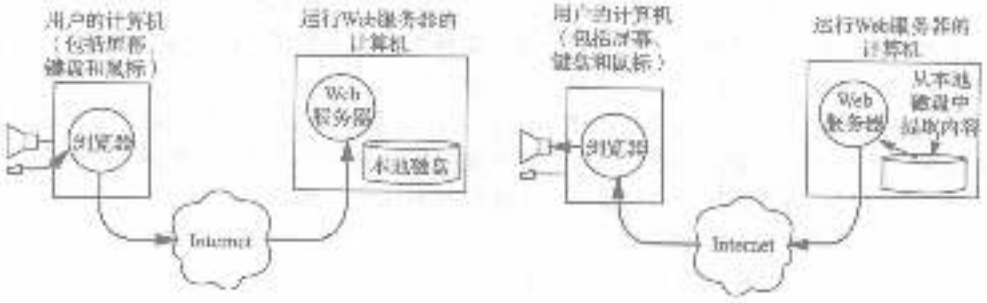
从理论上讲，任何一台计算机都可以作为服务器，大多数网站还是选择一些处理速度较快的计算机专门用作Web服务器。而且大部分Web站点的计算机往往配置了大容量硬盘来存储很大的数字图像。

因为Internet覆盖整个全球，所以网站在任何时刻都有可能被访问，这就要求网站提供不间断的服务（即，要保持一天24小时在线）。总之，网站必须时刻准备处理来自世界各地的访问请求。

第三个条件是一种软件，它才是将一台普通计算机变为能够响应网页请求的Web服务器的核心所在。为了能够做到“有求必应”，该软件必须一直运行。但是与普通软件不同的是，Web

服务器就不需要用键盘鼠标进行输入，也不会在屏幕上显示任何信息。Web服务器通过Internet与浏览器进行通信，通信时使用的是一种只有浏览器才需要明白的语言。浏览器就是通过这种特殊的语言向Web服务器请求网页。

图25-1展示了浏览器和服务器之间的通信情况。



a) b)

图25-1 a) 用户请求一个URL，浏览器会通知Web服务器取得需要内容；
b) Web服务器从本机磁盘上找户需要的内容，并返回给浏览器

浏览器使用URL中的信息来确定需要与那台计算机进行通信。每一个URL都包括计算机名和存储在该计算机上的一个对象的名称。获得一个URL后，浏览器首先提取计算机的名称，并通过Internet与该计算机上的Web服务器通信，然后请求URL中指明的对象。

如果没收到浏览器的请求，Web服务器就什么也不做，也就是说Web服务器一直都在耐心的等待下一个请求的到来，如果没有浏览器向Web服务器发送请求，该服务器就会永远等下去。不过，当计算机上的Web服务器程序处于等待状态的时候，其他的程序还可以正常运行，这样即便Web服务器处于空闲状态，该计算机也不至于完全闲置。实际上，一台计算机上通常都会有多种服务器软件在同时运行，例如文件传输服务器往往也是Web服务器。尽管这样一来，多个服务器必须共享一台计算机，但是一般用作服务器的计算机的处理能力都已经足够强大，以至于服务器是否处于"共享"状态根本不会影响到用户的响应时间。

首先浏览器与Web服务器建立连接并提出对某些内容的访问请求，然后服务器从请求中提取出所要访问的条目，并在本地磁盘中查找该条目，找到后给浏览器返回一个副本。

25.4 一次提取一项条目

一旦服务器将所需的条目返回给浏览器后，它们之间的连接就可以断开了。如果浏览器还需要从该服务器提取其他条目，它必须再次与Web服务器建立连接。大多数浏览器都会显示一些提示信息，让用户知道当前正在进行的操作，例如连接计算机、获取网页等。当浏览器正在获取的文件很大（如下载文件）时，用户希望知道还需要多长时间才能完成该操作。

因为大多数网页都包含许多条目，如果每次连接都只提取一项条目，Web服务的效率会很低。特别是对于包含很多图片的网页，因为每张图片在服务器上都存放在一个独立的文件中，每个文件都必须用一个单独的请求和响应来获取，如果浏览器请求的网页包含三张图片，它们之间就必须进行4次通信。第一轮通信过后，浏览器获得了主页面，因为本地还没有网页中包含的三张图片，所以暂时不能显示，必须再分三次请求这三张图片，然后才能显示给用户。为了克服这一缺点，最新的浏览器可以在获取页面所需各部分内容的过程中，始终与服务器保持连接。

25.5　传统的 Web 页面使用整个屏幕

普通页面会占据整个浏览器，当用户从一个页面跳转到另一个页面后，屏幕上的条目完全变为新页面的条目。如果用户还希望看到上一个页面的条目，他必须借助浏览器的后退功能。浏览器一次只能显示一个页面的条目，退回到前一个页面之后，就无法看到当前页面内容。

尽管浏览器这种一次只显示一个页面的方式能够满足大多数情况的需要，但还会有些不便。假如某个公司有六款主打产品，公司的网页上提供了这六款产品的列表，每一项都链接到介绍该产品详细信息的独立页面。浏览过程中当用户点击列表中的某款产品，浏览器就跳转到介绍该产品的页面。这种方式的缺点是，用户必须首先返回列表页面，然后才能进入介绍其他产品的页面。而且，当用户查看具体产品信息时，不能同时看到产品列表。

为了给用户提供更加便利的页面导航，网站的制作者往往在页面底部提供一些导航链接。例如，在每个产品页面的底部都提供两个链接：

返回公司主页

返回产品列表

25.6　Web 页面能改变屏幕的局部显示

为了能让用户更便捷的浏览列表信息，网页中引入了框架（frame）技术。使用框架能将浏览器的显示部分划分为多个矩形区域，并且一个区域内容的改变不会影响其他区域显示的内容。图 25.2 展示了浏览器如何显示使用了框架的网页。

如图所示，使用框架技术可以更方便的浏览列表信息。在图 25.2 中，页面左半部分包含了公司的产品列表，当用户选择并浏览一款具体产品的信息时，左边的列表仍然显示，并允许用户直接选择其他产品。图中显示的是用户选择"Toaster"后的页面显示情况。总结起来：

> 框架将浏览器的显示部分划分为多个可以独立控制的矩形区域，一个区域中内容的改变不会影响其他区域的内容。

图 25.2　使用框架技术的网页示例。在本例中选择列表中的不同条目后，左边的页面不会改变，而右边的页面会随之改变

25.7　Web、广告和框架

公司建立网站的初衷不是为了提供公共服务，而是具有重要的商业目的：提高公司的品牌形象以及产品与服务的知名度。简而言之，企业网站的主要目的可以总结为两个字：广告。企业网站上提供的信息可以吸引潜在客户或指导用户更好地使用其产品。

因为广告效应是企业网站的重要功能，为了能更好的发挥其宣传作用，企业开发了很多用于在网页上发布广告的技术。有些公司网站在每一个网页中都加入了广告信息，让所有访问公司网站的用户都能看到。还有一些公司设置了包含广告的中间页面，当用户点击链接查看感兴趣的信息时，浏览器会首先跳转到一个中间页面，该页面包含了广告内容和指向真正所需页面的

链接，用户需要再次点击链接才能到达所需页面。

很多网站也将框架技术用于做广告。公司网站的每个页面都可以使用框架，与上一节的例子所不同的是，这些框架并不都是用来显示相关的信息，而是选择一个固定的框架来显示广告信息。通常，广告所在的框架包含很多图片，颜色也更醒目，这样更容易引起用户注意。

在单独的框架中显示广告有利有弊。将广告置于页面的某个框架之中意味着它可以更长时间的出现在用户的屏幕上，优点是增加了广告被用户关注的概率；而缺点是企业不能将所有不同类的商品和服务做成广告。另外，使用框架也固定了广告在屏幕上的显示位置，这样做的优点是降低了对用户正常浏览的干扰，从而避免用户产生厌烦情绪；但缺点是容易被用户直接忽略。

当然，公司并不需要拘泥于某一种广告形式，可以通过框架机制将几种不同的广告形式结合起来。例如，公司可以将最重要的广告放到一个框架中，在其他页面中包含另一些产品，或者在一个小框架中提示用户如何获取更多的产品信息，而包含广告的框架中可能只包含下面一行内容：

点击这里查看更多产品信息。

框架中只包含一条简单的信息可以将屏幕上更多的空间留给其他框架。

25.8 弹出窗口与弹出窗口阻止程序

除了用框架显示广告之外，有些网站还使用弹出窗口（pop-up）技术。从名称就不难理解，弹出窗口是在用户没有请求的情况下，自动出现的新窗口。尽管弹出窗口可以显示一些有用信息（例如，推荐一些相关的网页信息），但更多情况下弹出的是广告，通常这些不速之客会影响用户的正在浏览的网页，因此不得不手动关闭它们。

因为弹出广告十分烦人，有些浏览器提供了阻止弹出窗口的功能，而且还有很多第三方的弹出窗口阻止程序（pop-up blocker）。

要点是：

> 网页中可以包含一些特殊内容，它会让浏览器自动弹出一些新的窗口。一般弹出窗口中内容都是烦人的广告，因此，可以用浏览器或其他软件阻止这类窗口的弹出。

25.9 静态页面的缺点

网页技术的设计初衷是用来储存那些在相当长时间内保持稳定的信息，就好像图书馆就是用来保存更新频率很低的书籍一样。事实上，那些研究电子信息检索技术的研究人员常常用数字图书馆（digital library）来描述他们所做的工作，所以早期网页技术主要是解决如何访问那些内容不变的文档。

尽管很多新的网页技术可以让人们更好的控制页面上显示的内容，但静态网页仍然流行，其原因有三：第一，静态网页容易创建。第二，静态网页维护费用低。第三，静态网页访问速度快。这些优点带来了很深远的影响：Web上的许多文档是静态的。

虽然静态网页如此流行，但它的缺点也不可忽视。首先，静态网页只能包含文本和一些图片信息，而不能包含其他形式的内容，如声音。其次，静态页面不能提供即时信息（如当前的天气情况和即时股票价格），因为所有的页面都必须事先准备好，如果有新消息需要发布，必须先用一段时间制作页面，然后用户才能看到更新的内容。第二，信息只能单向流动（从Web服务器到用户的浏览器），用户没有交互的体验。尽管静态页面也可以显示一张列表让用户去选择，但它不允许用户直接输入并搜索自己感兴趣的内容。第四，静态页面一旦加载并显示，不会再有任

何变化，用户必须进行某些操作（如输入新的 URL 或点击某个页面链接）才能让浏览器显示新的页面内容，因此，静态页面不能及时反映信息的变化，不能包含动画，不能自动跳转。

25.10　控制浏览处理数据的过程

为了克服静态页面的缺点，人们发明了很多新的网页技术，最初尝试的成果是让网页在文本和图像之外，还能够包含一段音频剪辑（audio clip）。音频剪辑是用户通过麦克风录制的一段声音。这段声音可以是音乐、对话或任何声音。而且声音的长短不限，可以只是一个响声，也可以是一段长篇演讲。

因为用户处理声音的方式与处理图像的方式不同，所以浏览器也必须将二者区别对待。如果网页中含有图片，那它一定包含一些说明信息，告诉浏览器那些图片需要显示，每个图片显示位置是什么等。浏览器根据这些信息自动获取图片并将其显示在适当的位置上。这样用户就可以看到图片又可以看到相关的说明信息。如果浏览器用同样的方式（自动下载、自动播放）来处理音频文件，就很容易让用户感到困惑，因为用户根本无从知晓这些声音的目的和作用。

为了让用户能够理解网页所包含的声音，就不能简单地将音频文件放在网页中，而必须要为每段剪辑都配上一个含有说明文字的链接。用户想听哪个内容就点击哪个链接。然后由浏览器获取音频文件，并通过音响或耳机播放。对于比较小的音频（像小到只有几个字的），浏览器可以在本机上保留一个副本，这样下次用户想再听一遍这段声音时就可以立即播放。总结的来说：

> 网页可以包含声音。当用户选择指向一段音频文件的超级链接时，浏览器会通过连接到计算机上的扬声器或耳机自动播放它。

看了上面的介绍，也许很多人会认为在网页中添加声音是很容易的事情。但是还有两个问题让事情并不像看起来那么简单。其中最主要的问题是音频文件的格式，因为还没有一个统一的标准，所以存在多种音频格式，但又没有哪一种格式可以胜任所有的需要。例如，可以选择用于保存 CD 音乐的音频文件格式，这种格式为了保证音乐的品质，占用了大量的存储空间，如果网页中包含的是这种格式的声音，那么浏览器就需要花更多的时间来下载，用户也就需要等待更长的时间。尽管还有一种占用空间更少更经济的音频格式可以很好的保存人们说话的声音，但是，采用这种格式录制的音乐就不会那么悦耳动听了。

除了音频格式的问题，为 Web 设计音频格式的设计者还需要将计算机的声卡也考虑在内。对于声卡同样没有一个行业标准，有些声卡支持播放立体声，有些就不支持，而且很多声卡都只能在特定型号的计算机和操作系统上使用，因此不可能保证一段音频在所有计算机上都能播放。

25.11　使用插件实现更多功能

就像网页中的声音具有多种格式一样，各种用于表示图像或文字的新文件格式也层出不穷。尽管可以通过不停的更新浏览器来支持所有这些格式，但浏览器厂商采用了一种更聪明的解决方法：在保证浏览器能够处理几种最基本格式的同时，利用某种技术让用户可以自行扩展浏览器可识别格式的范围。

这种技术就是插件（plugin）机制，它很好地解决了新格式所带来的问题。所谓插件就是一段专用的程序，它可以解释某种特定的数据格式。当某个公司或个人开发一种新格式时，他们还需要提供能够处理这种格式的插件。例如，针对某种音频剪辑的插件必须能够操作硬件来播放声音。如果用户想要使用这种格式，就要下载这种插件来扩展浏览器。

很多浏览器在初装时就包含了一些常用的插件，而且安装新插件也很容易。更重要的是，网

页可以自动检测用户浏览器是否含有某个需要的插件，如果没有，就通知用户，并自动下载安装。用户在正常浏览各种网页内容时，如果选择了需要特定插件才能处理的内容（如音频），浏览器会自动检测该插件是否存在，如果已安装就调用该插件来处理这段内容；如果没有安装就通知用户。通常情况下，浏览器会在屏幕上显示所需插件的名称，下载插件的地址等信息，并等待用户来采取下一步操作。用户可以下载并安装插件，也可以不安装插件，这样浏览器会忽略这段不能解释的内容[1]。如果用户决定安装，插件会被下载并安装到本机上的浏览器中，以后每次启动浏览器，都会自动加载这些插件，这样用户就可以利用该插件自动处理需要的内容。

25.12 服务器可以按需生成页面

尽管插件机制可以处理多种类型的数据，但是这还不足以弥补静态网页的所有缺陷。所以工程师们又开发出一套被称为服务器端脚本的技术（server-side scripting technology），它可以让网页包含动态内容[2]。这类动态网页不像静态网页那样只是在服务器上保存一个不变的副本，而是在被请求时才去生成页面所需的内容。静态网页时代所使用的URL必须指向服务器磁盘上的某个静态文档，而服务器端脚本则可以将URL关联到服务器上一个可执行的程序。当浏览器请求这类URL时，服务器首先运行对应的程序，然后将程序的输出结果返回给浏览器。服务器上可以保存任意多个脚本程序，然后根据用户的URL请求选择不同的程序来完成所需的计算任务。

为什么用程序输出的内容比原先的静态文档更具优势呢？因为程序可以获得更多数据并完成更复杂的计算。例如，Web服务器可以和一个电子温度计相连，这样就能利用特定程序动态采集温度数据，并以网页形式返回给用户浏览器。如果Web服务器能够利用某些程序即时获取股票价格信息（这里的“即时”是指用户请求网页时的价格，而不是当天的开盘价），它同样可以将这些结果生成网页返回给用户。类似地，网上商店可以利用服务器端脚本技术为用户提供包含准确的库存信息的商品目录（这里的“准确”是指用户请求网页时的数量，而不是昨天盘点时的库存量）。又如果Web服务器与一个摄像头相连，它就可以获取一张实时图片放入给用户的应答网页中，这样就构成了一个简易的监控系统。

实时计算能力并不是服务器端脚本技术的惟一特性，程序还可以将某些信息存储在磁盘上，这样就具备了保留历史记录的能力。最简单的应用莫过于记录某个网页被访问的次数：使用一个计数程序将每个网页的访问次数保存在服务器上的文件中，每当用户访问一个网页，计数程序就增加相应网页的计数，然后将网页的HTML内容连同计数值一并返回给用户。尽管这个过程看上去有些复杂，但对于程序员来说则是小菜一碟（几行代码而已），这样就实现了一个简单的计数器。要使用该计数器，可以在网页中加入这样一句话：

网页已经被浏览N次。

其中N代表网页被访问的次数，每当有用户访问该网页，计数值就会自动累加，当然用户刷新页面也会使计数值增加（因为计数程序无法区分新的请求和刷新请求）。

25.13 服务器端脚本如何工作

那服务器端脚本是怎样工作的呢？简单起见，我们可以将脚本看成Web服务器的一部分。浏览器向Web服务器发送请求，如果用户请求的URL对应的是一段脚本，服务器就运行相应的

---

[1] 有些插件只提供了针对某种浏览器或计算机的可执行版本。

[2] 最早的服务器端脚本技术称为CGI（Common Gateway Interface），也称作CGI-bin。

程序，其中请求的 URL 地址会作为一个参数传递给该程序。然后服务器就等待程序的输出结果，并把该结果作为对用户请求的响应返回给浏览器。图 25-3 阐明了这个过程。

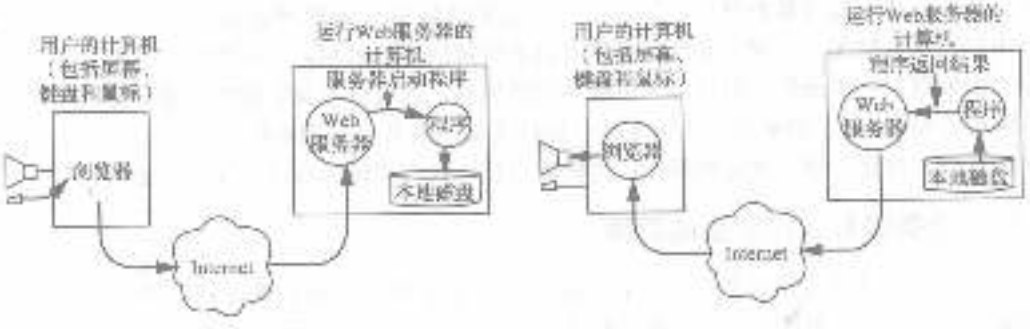


图 25-3　a) 浏览器请求的 URL 对应着一段程序，因此服务器会启动这段程序；
b) Web 服务器最后将程序运行的结果返回给浏览器

有意思的是，从浏览器的角度看上去，无论请求的 URL 指向的是静态网页还是动态脚本，访问的结果都没有差别。而且访问这两种不同类型页面的请求格式也完全相同。类似的，服务器对这两类请求的应答也使用相同的格式，因此，浏览器根本无法区分从服务器获得内容是静态网页的一份拷贝还是动态脚本运行的结果。

25.14　程序员创建服务器端脚本

由谁来创建脚本呢？只有具备计算机程序设计能力的程序员才能创建服务器端脚本程序。因此，大多数普通用户都无法胜任。更重要的是，脚本程序有别于普通计算机程序，它们直接与 Web 服务器进行交互。因此脚本必须符合 Web 服务器的要求。例如，脚本不能像普通程序那样接收键盘输入并将处理结果显示在用户屏幕上。相反，Web 服务器首先获取脚本的输出结果，然后再传输给用户浏览器。

有趣的是，Web 服务器和脚本的正常运行也和计算机所使用的操作系统息息相关。因此，和普通程序一样，完成某一功能的脚本程序只能在特定 Web 服务器软件和操作系统上运行。如果想在其他计算机系统上完成同样的功能，很可能需要重写这段脚本。

25.15　个性化 Web 页面

人们很快就意识到可以利用动态生成网页内容的技术来为不同的用户量身定做个性化的 Web 页面，从而提升 Web 应用的吸引力。过去只能先准备一系列静态网页，然后由用户来从中挑选自己想看的内容，而现在服务器可以直接生成更符合用户口味的页面。例如，服务器可以自动调整网页内容的显示顺序（将用户最感兴趣的内容放在显眼的位置）。甚至还可以根据所掌握的用户信息来选择不同的页面内容。

为了更好的理解个性化 Web 页面的优缺点，我们来分析一个提供个性化歌曲列表的页面。假设服务器可以通过某种方法近似的获知用户的年龄层次，这样，如果访问该页面的是青年人，服务器会选择当前最流行的歌曲；如果是中年人，服务器则会挑选 70 年代的风靡一时的摇滚歌曲；如果是 70 岁以上的老人，页面返回的结果就是 50 年代的老歌。总之页面的内容尽可能满足不同用户的个人口味。

然而，个性化的内容也有一个严重的问题：页面内容可能无法重现，访问同一个页面的两个

用户很可能看到不同的结果。有些情况下这的确会成为一个问题。例如，一个中年母亲想要给自己的孩子购买一些音像制品作为礼物，恰巧，她的孩子是通过个性化歌曲列表页面来选择他自己喜欢的内容，并让他的母亲通过访问网页的URL来查看自己选择的结果。由于该网页具有个性化特性，母亲所看到的歌曲列表很可能和她的孩子选择的结果大相径庭。要点是：

> 利用动态内容技术，服务器可以动态地决定网页的内容和显示的顺序，从而创建个性化网页。尽管个性化页面能够照顾用户的兴趣所在，但它也可能带来页面内容无法重现的问题。

25.16 个性化广告

前面介绍过的框架技术可以将广告内容和正常网页内容分离。服务器上的脚本可以记录用户的浏览历史情况，并据此选择合适的广告信息。整个过程如下：当服务器收到用户的页面请求，首先会运行脚本选择一个合适的广告，然后将广告和所请求的网页分别放在不同的框架中返回给用户的浏览器。如此一来，每次用户访问网站，服务器都可以返回不同的广告内容。

服务器上的脚本还可以根据用户访问网站类型的记录来选择合适的广告。如果用户要浏览有关家具和电器的网页，服务器可以选择关于家装家居方面的广告。如果用户要浏览有关流行音乐的网页，服务器就可以选择音乐会门票之类的广告信息。

25.17 Web页面也可以交互

尽管前面介绍的脚本技术可以用来生成具有动态内容的网页，但是它们只能在服务器一端运行。因此，用户并不能直接通过脚本实现交换。如果用户想要输入数据，就必须使用一种称为表单（Form）的技术。包含表单的页面中会出现一些空白区域，用户在这些指定区域输入信息后，浏览器可以将它们发送到服务器上。

表单的优点很明显：让用户可以自由的输入信息，而不再局限于从已有的内容列表中去选择。例如，当用户在网上购买商品时，他必须提供信用卡号和收货地址，没有表单技术，输入这类信息几乎是不可能的。采用表单之后，网页中会出现让用户输入卡号和地址信息的空白区域。用户输入完毕后，只需要再轻轻一点（例如，带有“购买”字样的按钮），浏览器就会将这些信息发送到服务器一端。

25.18 购物车

购物车（shopping cart）是一类十分有用的个性化页面。正如它的名字一样，这种页面模仿了现实生活中超市所使用的购物车。以销售商品为目的的商业网站会使用购物车。在浏览网站的过程中，用户随时可以创建购物车。当用户看到需要的商品时，可以点击标有“加入购物车”字样的按钮或超链接，服务器会保存该用户购物车中所有的商品列表，并允许用户随时查看该列表。

购物车不会长时间地存在，通常，它们的有效期只有几个小时，最多也就是几天时间。如果用户购物的时间超出了这个期限，购物车就会因超时而失效，此时用户就必须再创建一个新的购物车并重新挑选商品。当然，大多数用户都会在购物车失效之前“结账”（提交订单）。

25.19 cookie

服务器是怎样记住每个购物车和用户的呢？它又是如何判断某个用户是否曾经访问过某个网页的呢？很多系统都通过这样一种方法：为每个用户和订单分配一个编号。用户可能会看到这

个编号，也可能根本不知道编号的存在。例如，某个网站可以将订单编号通知用户，并要求用户下次访问时输入这个编号。但是大多数情况，这个特殊的编号是对用户隐藏的。一种巧妙的实现方式是：服务器在所发送的每个页面内都嵌入编号信息。而用户在每个页面请求或应答中也偷偷地包含同样的编号。

值得注意的是订单编号并不包括购物车中所包含的具体商品信息。相反，它仅仅是一个标识符，让服务器可以根据该编号找到对应的购物车，这也就是说服务器必须负责保存每辆购物车中的商品列表。用一个更短的标识符来替代一长串实际内容是一种很有价值的技术，它可以减少浏览器和服务器之间交换信息的数量，大大加快通信的响应速度。

对于一张订单来说一个购物车编号就足够了，但如果某个用户在一个月之内两次访问某网站，并提交了两个订单，那么服务器必须为他或她分配两个不同的购物车编号。

那么服务器怎样发现同一个用户多次访问网站的情况呢？最简单的解决办法是利用并扩展购物车编号的方法，为每个用户也分配一个特定的编号。为了能让这一机制对用户透明，浏览器与服务器需要紧密配合。当用户第一次访问某个网站时，服务器为用户分配一个独一无二的编号，并把这个编号返回给浏览器。浏览器会将这个编号保存在用户计算机的磁盘上。当用户再次访问该网站时，浏览器会自动将用户编号发送给服务器。

浏览器和服务器用特定编号来标识不同用户，这个标识还有着一个奇特的名字：小甜点（cookie）。和购物车的标识符类似，cookie 体积也很小，只保存了部分用户信息。更详细的用户信息都保存在服务器的磁盘上，cookie 只是作为索引来方便查找。因此，尽管 cookie 很小，但服务器可以保存任何与用户相关的信息（例如，该用户浏览过哪些网页，向该用户投放过哪些广告，用户每次访问都买了哪些商品等）。

25.20　应该禁用 cookie 吗

任何时候，服务器都可以向浏览器发送 cookie。但是在浏览器接受之前，它会询问用户怎样处理收到的 cookie。很多浏览器都允许用户选择不同的处理方式，可以拒绝所有的 cookie，也可以接受所有的 cookie，还可以在每次收到 cookie 时都提示用户来处理。因此，用户必须决定处理 cookie 的方法。

接受 cookie 的好处是它可以让服务器利用你的浏览历史，从而更好的提供个性化的内容，但 cookie 的缺点是向服务器暴露了你的浏览和购物习惯。总结起来：

> cookie 是浏览器和服务器之间用来标识用户的一个值，接受 cookie 可以让服务器针对用户的兴趣爱好来提供网页和广告内容，而禁用 cookie 可以保证隐私性。

25.21　Web 页面中也可以播放动画

自从 Web 一问世就有很多人问："为什么网页不能像电视一样显示连续的画面？"答案是，Web 技术和 Internet 最初都是设计用来处理静态文档，而不是传输和播放视频。为了获得动画效果，计算机必须快速显示一系列图片，即使服务器端脚本也不能生成图片流，脚本的输出结果是服务器用来应答浏览器的一个具有固定内容的文档。

现在有些技术可以让网页包含动画，其中最早的是一种被称为客户端拖拽（client-pull）的技术。之所以用这样一个略显滑稽的名字是因为它要求浏览器反复不断地向服务器发出图片请求。通过这样的过程，让浏览器从服务器上"拖"下来一系列图片，然后再顺序播放它们。

那么浏览器又是怎样知道它需要进行拖拽的呢？答案就存储在被访问的网页中。创建支持

客户端拖拽的网页序列时，作者会在每个网页中加入一些特殊的标记，这些标记指明了该页面所需的后续页面，以及两次请求之间的时间间隔。这样网页的作者可以安排网页的显示顺序：第一个页面会命令浏览器去请求第二个页面，第二个页面又会命令浏览器去请求第三个页面，依此类推。为了让这些页面连续播放，最后一个页面会让浏览器重新请求第一个页面。

从理论上说，客户端拖拽技术可以在用户屏幕上产生动画效果。因为，电视也不过是一种更新频率更快的图片序列而已。如果将网页序列中的每个页面都看作电视画面中的每一帧，并且将两次请求的时间间隔设定为零，浏览器就会连续请求页面，用户就可以体验动画效果。

然而，这种设想在实际中行不通。最大的问题就在于客户端拖拽技术所产生的延迟太大。为了让人产生动画效果，浏览器每秒至少要显示 24 张图片。即使在一台拥有高速 Internet 链接的高速计算机上，浏览器也需要很长时间来获取并显示一张图片，根本不可能达到产生动画效果的速度。而且，Internet 上数据传输的速度也不稳定，这会导致动画播放速度时快时慢，给人一种时断时续，画面跳跃的感觉。因此，客户端拖拽技术只能显示一些很简单的动画（例如，一个人前后挥动手臂）。实际上，浏览器往往因为等待后面的图片而在当前的画面上停顿几秒钟。

25.22 活动页面功能更强大

为了更好的播放动画，人们又开发出了一些更新、更强大的 Web 技术。这些新技术采用了一种完全不同的方法来实现活动（active）网页。活动网页是一段能直接在屏幕上输出图形图像的程序。与普通网页一样，活动网页也保存在 Web 服务器上，并与一个 URL 相关联。而且，访问方法也一样：浏览器用 URL 来请求某个活动页面，Web 服务器返回该页面的拷贝。但是浏览器处理活动页面的方法和普通网页不同，它首先在本机上运行页面中的程序，由程序决定浏览器中显示的内容。图 25-4 阐述了这个概念。

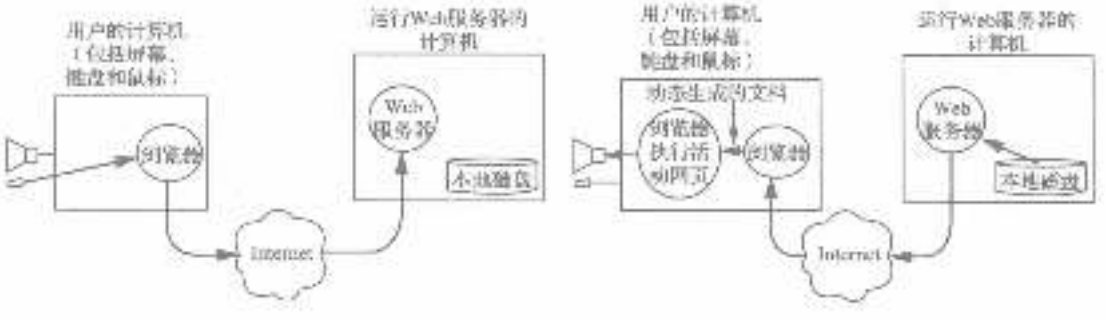


图 25-4 a) 浏览器请求的 URL 对应着一个活动文档；
b) 浏览器执行活动文档的拷贝，并自此控制屏幕输出

25.23 Java 是一种活动页面技术

Sun 公司（Sun Microsystems，Inc）开发了一种广受欢迎的活动文档技术——Java。这项技术可以创建能与用户交互的动画网页，也可以让网页以一种意想不到的方式显示在屏幕上。在 Java 中，一个活动网页又被称为小程序（applet），该术语非常流行，其他公司也纷纷采用小程序这个名字，或者使用一个与之类似的名称。

Java 的流行主要有四个方面原因。第一，Java 语言在设计上借鉴了其他流行的程序设计语言，这意味着程序员可以很容易学会编写 Java 小程序。第二，它是第一个用于活动网页开发的

语言。第三，Java 语言提供了功能丰富的函数库，能够完成诸如绘制屏幕等常见任务，程序员可以利用这些现有的功能模块快速完成所需小程序的开发。第四，Java 语言比其他技术提供了更多、更强大的功能。例如，Java 可以比表单更好的实现用户交互，比客户端拖拉更好的获取一系列页面，比框架更好的分割屏幕区域，而且比插件更好的操作各类数据。因此，可以说 Java 集各种功能于一身。

尽管与现有的技术相比具有很多优势，但 Java 成功的最大驱动力还是源自于它所提供的一种独一无二的功能：高品质动画。因为 Java 是利用客户端计算机的处理能力来即时生成图像，而不再是从服务器上下载，所以用 Java 制作的活动网页可以快速的改变屏幕显示让用户看到流畅的动画效果。因为其他技术都无法做到这一点，所以很多网站都争相使用 Java 技术。

25.24　JavaScript 脚本也是活动页面技术

尽管 Java 是最强大的活动网页技术，但它也是最复杂的一个。编写可以控制屏幕图像的 Java 程序需要编程人员具备深厚的基础知识并经过专业的训练，因此它只适用于高级程序员。

随后，一些 Java 技术的替代者相继问世，JavaScript 就是其中最受欢迎的一个。一方面它保留了很多 Java 的基本特性，另一方面又去除了 Java 中很多复杂的部分。JavaScript 可以嵌入到标准的 HTML 文档中，当浏览器看到 HTML 文档中的 JavaScript 部分时，便执行其中的计算任务，并返回结果。如此一来，JavaScript 用一种相对简单的方式完成了 Java 可以完成的绝大部分工作。

25.25　Flash 动画和 Real 流媒体技术

这两者是目前十分流行的，能够为活动的网页增加动画等视觉特效的技术。Macromedia 公司（已被 Adobe 公司收购）开发了 Flash 技术，它是一种专门用于网页的动画技术。RealNetwork 公司开发了 RealAudio 和 RealVideo 技术，它们可以独立使用，允许浏览器分别播放音频和视频。以 Java 和 JavaScript 脚本为代表的程序技术和以 Flash 和 RealVideo 为代表的多媒体技术的主要区别体现在，两者协调图像和声音的方式不同。多媒体技术可以让设计者控制音频和视频的细节，并实现与具体计算机无关的特性。例如，设计人员可以精确控制视频剪辑播放的长度，而不必担心剪辑播放速度因计算机处理器速度不同而不同。类似地，设计者还可以控制音频和视频协同播放，要点是：

> 现在有很多实现活动网页的技术，除了 Java 和 JavaScript 等编程语言之外，还有一些多媒体技术可以为网页添加视觉特效、音频和视频等效果。

25.26　高级 Web 技术的重要作用

目前 Internet 上还有很多静态网页，浏览这类网页与从数据库中检索数据的过程很类似，这些网页信息稳定，很少改变。但是，利用本章介绍的这些技术，可以制作动态网页，动态网页会根据浏览者的兴趣爱好显示不同的内容，也可以将一个页面分成多个独立的部分，还能通过对话框来运行用户进行直接的交互。

Java、JavaScript、Flash 和 RealVideo 等动态网页技术变得越来越流行，并逐步取代了传统的服务器端技术。带来这种变化的原因主要有两个。首先是经济原因：服务器的所有者不可能依靠不断升级硬件来满足所有用户的计算要求。第二个是功能原因：动态网页中的元素具有更好的显示效果，流畅的动画和视频已经极大地改变了网页的面貌。

第 26 章　群组和个人网页（维基和博客）

26.1　引言

前几章我们介绍了电子邮件（email）、公告板（BBS）和网页（Web）在 Internet 中的使用。本章我们继续讨论两种可以让群体或个人在 Internet 上创建网页的新服务。

26.2　论坛的缺点

我们在第 22 章讨论了新闻组，它是一种公告板服务，允许订阅者发布和浏览文章。可以说新闻组给每个人提供了一种可以针对任意话题各抒己见的机会。新闻组可以涵盖所有主题，每当有一个新的主题出现，都需要创建新的组。

尽管新闻组能够让很多用户表达观点，但它和实际的公告板具有同样的缺陷。例如，如果管理不善，任凭新的公告随意覆盖旧的公告，那么最终公告板将变得一团糟以至于根本无法使用。新闻组也有类似的问题：如果各类文章鱼龙混杂，会影响用户获得有用信息。

新闻组与实际公告板的相似之处可以归结为两点：第一，绝大多数新闻组允许用户随意订阅或发布文章，这样很容易造成讨论内容偏离原先的主题。第二，文章是根据发布时间进行排序。为了理解跑题问题，我们举一个例子。假设有一个讨论足球的新闻组，内容包括球队、球员、比赛、战术等。然而，一个不恰当的评论甚至用词都可能将讨论引入歧途。像带有政治色彩的词"共和党的球队"，或一句抱怨"门票价格太离谱了"，都可能引起一番热评，但是它们都与新闻组的主题无关。

这种组织方式之所以会产生问题有两个原因。首先，文章列表中包含很多与主题无关的讨论，这样读者必须自行跳过这些无用文章。第二，因为文章的出现的顺序由发布时间来决定，这容易造成不必要的重复讨论，让读者产生困惑。假设新闻组中的某个主题有 6 条消息：

$$A,\ B,\ C,\ D,\ E,\ F$$

刚刚度假回来的读者 Joe，首先会看到序列中的第一条消息，并作出回复。对于其他读者来说，这条回复是一条新的消息 G，这时 Joe 只读完了消息 A。但是当 Joe 看了从 B 到 F 的讨论后，感觉到刚才自己的回复根本没有必要，他又发了一条消息 H 表示歉意并建议其他人忽略刚才的回复的消息 G。经过这一系列操作后，该主题下已经变成 8 条消息（很显然最后两条完全是多余的）：

$$A,\ B,\ C,\ D,\ E,\ F,\ G,\ H$$

在一个拥有数以千计的读者的新闻组中，各种文章的顺序会更加凌乱。假设对一个新来的读者 Nancy 来说，她需要花费很长时间才能理清头绪。

26.3　共享页面

除了论坛之外，还有一种共享页面服务（page sharing service）。与 BBS 不同，共享页面服务不会将各个回复罗列出来，也不保存每个回复的历史记录，它仅仅创建一个用户共享的页面。每个读者都可以修改它的内容，而且修改结果对所有用户都可见。因此，当一个用户替换了一段内容，所有用户都会看到更新，而且不能再恢复原文。

尽管不同的共享页面系统有一些差异，但大多数都是匿名（anonymous）的。也就是说，系统并不会记录那些编辑了页面内容的用户的姓名或信息。所以对读者来说，不可能知道页面的内容是由谁提供的。

26.4　共享页面就是维基

Wikimedia 基金会（Wikimedia Foundation）是共享页面技术的开拓者。现在共享页面有了一个更流行的名字，维基（wiki），而实现共享页面的软件被称为维基软件（wiki software）。维基这个词是从夏威夷土著语"wiki wiki"而来，它的意思是"快点，快点"。

这个名字起得十分恰当，wiki 软件不用等待页面被彻底更新，只要对 wiki 中的内容进行编辑后，改变会立刻完成。

总结起来：

> 维基是一个由很多用户所共享的网页，某个用户修改其内容后，其他用户能够立刻看到新的内容。

26.5　共享与共识

共享页面的主要初衷是建立共识（consensus building）——让一群人可以迅速达成一致，然后再修改细节。为了阐释 wiki 在这方面的优越性，我们将 wiki 与新闻组做一下简单比较。如果一个 6 人小组通过新闻组或邮件列表的方式达成一致，讨论过程中所产生的各个消息只是按照时间顺序排列，缺少上下文关联；如果换成 wiki，那么整个讨论过程中每一次的修改意见以及最终的结果都一目了然。很明显，使用 wiki 可以避免误解，迅速形成统一意见。

26.6　维基的缺点

尽管通过 wiki 可以更轻松地获得统一意见，但是当大家的意见相左时（不同的声音很多），不但不能加速建立共识，反而造成不同意见泛滥。假设一个小组用 wiki 技术建立了 Web 共享页面来进行意见交互，如果某个用户不同意其他用户提交的内容，他可以自由地进行修改。即使持反对意见的人只有一个，他也可以通过加入一个否定词（not），让句子表达相反的意思；如果最初发表意见的人发现内容被篡改并去掉了否定词，他还是可以再改回去。

要点是：

> 虽然共享页面技术有助于多人达成共识，但因为它允许每个用户对内容进行任意修改，所以一旦出现持不同意见者，反而会阻碍统一意见的形成。

26.7　维基百科是一个试验

Wikimedia 基金会正在使用共享页面技术进行一次大规模的试验，试验的内容是创建一个名为 Wikipedia 的网上百科全书，通常也称为维基百科。其内容完全由用户提供或修改。维基百科是免费的，任何人都可以使用它查找信息。

由于人们匿名走进入维基百科，所以维基百科中的条目一般都是匿名的，当然，普通的百科全书也不会标明每个条目的作者[⊖]，但二者的差别在于，维基百科的作者具有不可追踪性，即使维基百科的所有者也不知道很多条目的作者。总结起来：

> 维基百科是一个基于 wiki 技术的、免费的在线百科全书。任何人都可以匿名

⊖ 这些内容的作者曾经给出名的百科全书写过文章。

为其添加新的内容或修改已有内容。

26.8　维基百科是否可信

维基百科上的大多数条目都是由善意而理性的作者提供，他们尽可能的提供真实而有价值的信息。然而，匿名总会招致一些问题，有些错误内容是由于疏忽大意造成的，但有些却是恶意为之。因为维基百科的所有者没有控制编辑权限，不去挑选作者、不检查提交的内容、不要求作者提供身份信息，所以用户必须小心判断内容的可信程度。事实上，有些内容是由专家提供的，而有些则是由业余爱好者或初学者所添加。

由于新闻组和 wiki 建立在公平开放的原则上，缺少权威的控制手段，这导致一些 Internet 用户质疑其内容的可信性。在 2005 年 11 月的一份今日美国报（USA Today）上就报道了这样一个例子。维基百科上的一篇有关 Seigenthaler 的传记声称他与肯尼迪（John and Robert Kennedy）遇刺有关。这篇缺乏可信度的文章被其他网站转载，会误导读者。

处理 Internet 上的问题总是很困难。美国的隐私法禁止网站的所有者在未经提供内容的作者同意的情况下泄露其身份信息，除非作者授权或在法庭上得到允许。而且，现有法律使得电子媒介公司可以免予遭受诽谤诉讼。尽管最后 Seigenthaler 将虚假内容从维基百科上删除，但这件事给我们提出一个警告：

> 因为维基百科允许匿名提交内容，而且不用保证内容的正确性和准确性，所以很可能包含误导他人的错误信息。

26.9　个人日记的公开发布

除了以匿名方式产生信息的 wiki 以外，还有另一种 Internet 服务允许用户匿名发布信息，发布自己的所见所闻，观点和评论等内容。因为与普通的日记类似，所以被称为网络日志（Web log），也称为博客（blog）。利用这种技术可以让用户轻松创建包含各种内容的网页。

博客的题材内容广泛。有些主要记录与作者紧密相关的信息，读者也主要是作者的家人和朋友；但也有些博客涉及的内容更宽泛，如对时事新闻的评论、社会动态、政局等的评论。

博客和维基的差别主要体现在两个地方。第一，博客通常由一个人完成。第二，博客不是匿名的——作者一般会保证内容的真实性。简而言之，尽管博客上的内容体现的是一种个人意见，但可以保证消息的来源相对固定，因此便于人们了解作者情况，并根据作者的背景判断博客内容的价值的大小。

26.10　个人见解

在 Internet 还不像现在这样流行时，我写了几篇关于计算机科学的诙谐散文，并共享给我的学生们。起先，我将文章连同一些我写的小程序放在实验用的 FTP 站点上，当万维网兴起时，我便把这些散文移到了个人主页上。

前几年，当我访问某所大学时，有一个研究生兴冲冲地跑到我面前询问我在写那些散文时是否参考了他的博客中的文字。起初，我认为他在谈论我的博客，就解释说我没还没创建自己的博客，后来他解释说他所指的是我个人主页上的那些散文。也许是我文章中的某个用词和他的博客中某篇文章的标题类似，而且他希望这不仅仅是一个巧合。我很肯定地告诉他，我的那些文章都是在他上大学之前完成的，听到这些，他有些沮丧，连忙对我解释说他在半年前建立了一个博客，但好像根本没有读者。我很想给他打打气，可是我又能说什么呢？事实就是如此：Internet 的发展可以让人们更容易的制造信息，但却不能保证这些信息一定会有读者。

第27章　网页搜索（搜索引擎）

27.1　引言

我们在前面几章介绍了如何通过交互的方式在万维网上获得信息。浏览或导航（browsing or navigating）过程需要用户提供网站地址。连接Internet之前，用户必须输入URL作为起始点。在后续浏览过程中，用户要么跟随页面中的连接，要么手工输入其他地址。

本章将介绍一种有别于指定URL的查找信息的方法：自动搜索服务。使用该服务，用户可以不需要事先知道包含所需信息的网页的URL。该服务还允许用户跳过烦琐的页面链接跟转，直接访问位于远程计算机上的所需页面。更重要的是，自动搜索服务可以根据用户给出的主题或关键词返回一系列相关网站。本章不仅会介绍搜索服务的工作原理，还会介绍它的用户界面。

27.2　搜索引擎的功能

自动搜索服务（automated search service），又被称为搜索引擎（search engine），允许用户查找存在于远程计算机上的网页信息。自动搜索与网页浏览和文件获取的差别在于，它既不需要用户指定远程计算机，也不需要跟踪多重链接。相反，自动搜索系统利用程序寻找与给定主题相关的网页。特别的，它可以自动定位下列网页：

- 包含特定主题信息的网页
- 与特定公司或个人相关的网页
- 包含特定产品或品牌信息的网页

为了方便用户访问，搜索结果以网页形式返回，而且找到的每个结果项都对应网页中的一个链接。这样，用户在标准的浏览器下，很容易根据搜索结果找到需要的信息。

27.3　浏览还是搜索

尽管直接浏览网页更富有乐趣，但是在现有的Internet规模下，用户已经不可能使用逐个检查每台计算机的方式寻找所需的信息。实际上，Internet的持续增长让这种搜索徒劳无功，新网页产生的速度远远大于用户浏览的速度。

为了适应Internet不断增长的需要，自动搜索服务应运而生。自动搜索利用程序自动访问Internet上的其他计算机，寻找特定信息，并生成结果报告。这类程序被为搜索工具（search tool），也称为索引工具（indexing tool）或搜索引擎（search engine），而它们提供的服务称为自动索引（automated index）。

27.4　搜索引擎可以更好地帮助用户入门

对于刚开始接触某个新主题的用户来说，自动搜索服务特别适用。如果考虑到浏览器的工作方式和Web的组织方式，这一点就不难理解了。浏览器只能显示用户指定的URL中的内容，因此，用户至少要提供一个初始URL。又因为Internet上的信息并不是分门别类地存在，用户并不能简单地猜测哪里是起始点。假如，有个用户想查看一些与汽车相关的信息，他可能知道一

其汽车厂商的名字，或者能写出其中某个公司主页的 URL 地址。但是能做到这一步用户又能有多少呢？只有少数用户会知道汽车厂商的名字，能猜中 URL 的用户就更加是凤毛麟角了。

面对于更一般的主题，猜中 URL 几乎是不可能完成的任务，而且，以 Internet 的现有规模和增长速度，也使得用户不可能用手工的方式进行地毯式的搜索。因而，需要解决的问题就从"人们应该从哪里开始查找需要的信息？"变为"针对某一个主题，人们怎样才能找到一些有价值的起始点？"事实上，Internet 上的信息浩如烟海，我们只能将注意力放在对我们有价值的地方。

搜索引擎能够提供这方面的帮助。当用户开始收集某一方面的信息时，他可以根据主题做一些关键字检索，搜索引擎会返回一个包含相关信息的页面的列表，用户再进一步查看这些页面中是否包含他感兴趣的内容。如果用户认为某个网页有价值，可以记下它的 URL 地址，否则就继续查看列表中的其他网页。要点是：

> Internet 本身是一个巨大的信息库，在其中寻找某些特定网页十分困难。自动搜索能够帮助用户找出一组与给定主题相关的网页。

27.5　搜索工具可以使用户避免信息丢失

自动搜索工具不仅能够帮助用户找出浏览的起始页，还可以帮助用户找回丢失的网页地址信息。我们举一个例子来说明搜索工具这方面的作用：假设有个用户花了几个小时在网上查找某类信息（例如，最新的电子游戏等），不巧的是，就在他刚刚找到一个想要的网页时，他的计算机因雷击而意外重启。[⊖]当计算机重新启动后，他重新打开浏览器，但是浏览器不会自动载入刚才的页面，而是载入用户之前设置的起始页面。为了找到重启之前的那个页面，他要么重复刚才的查找过程，要么使用搜索工具找出该页面的 URL。总结：

> 自动搜索工具可以帮助用户找回丢失的网页地址信息。

27.6　自动搜索服务是怎样运行的

理论上，当用户提出一个请求时，搜索引擎需要访问 Internet 上的所有计算机来寻找符合要求的文档。搜索引擎首先要判断一台计算机是否是 Web 服务器，如果是还需要进一步核查服务器上的所有网页。在现实中，要将 Internet 上的所有主机部搜索一遍恐怕要花费几周的时间，没有用户愿意等待这么长的时间。好在搜索引擎只需要几秒钟就可以返回结果。那么搜索引擎为何能够如此迅速的完成搜索呢？

为了理解搜索服务能够如此快速响应请求的原因，我们先来分析一下电话黄页的工作原理。为了方便用户高效地找到需要的电话号码，电话公司收集每个用户的姓名、号码等信息，并将这些信息根据字母顺序排列，形成一个巨大的电话簿。当有人想找某个城市中所有叫 Jones 的人时，他不需要把所有装电话的家庭问个遍，只需要找一本该城市的电话黄页，并查找其中以 J 开头的部分即可。

Internet 上的自动搜索服务采用类似同样的策略来加速查找过程。在用户提出搜索请求之前，搜索引擎会自动联系所有连接到 Internet 上的计算机，收集各种网页信息，对它们进行分类排序，并建立索引，最后将索引保存在本地磁盘上。要使用搜索服务，用户通过浏览器向搜索引擎发送一些搜索关键字。当搜索引擎收到搜索请求后，无需再查找 Internet 上的计算机，只需要搜索本地磁盘上的索引文件即可（与通过黄页查找电话号码的过程类似）。

⊖ 当用户突然拔掉信息或未记录网页地址就关闭浏览器时，会发生信息丢失的情况。

27.7　预先收集信息

预先收集信息确实可以大大加快搜索的速度，但也有一个弊端：不能及时反映页面的变化情况。电话公司解决这个问题的方式是：定期更新电话黄页，一般是一年一次。自动搜索系统的处理方式也类似：定期运行程序访问 Internet 上的所有计算机，检查内容更新情况，生产新的页面索引。例如，好的搜索引擎会利用蜘蛛（spider）程序不停的探测 Web 上的内容[⊖]。当蜘蛛程序发现新的条目后，会将其加入搜索引擎的数据库中，如果以后遇到用户对这方面内容的搜索，搜索引擎就可以将这些条目返回给用户。图 27-1 阐释了这个过程。

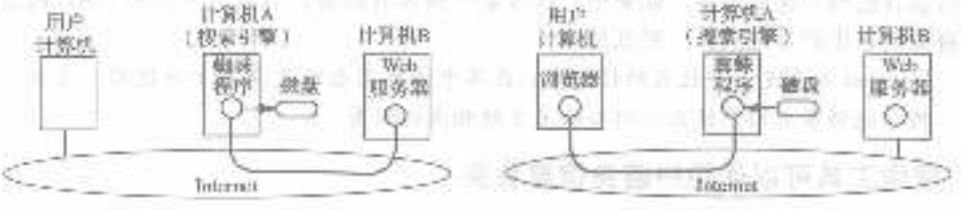


图 27-1　自动搜索服务所使用的两阶段过程的示意图。a）蜘蛛程序自动访问每个 Web 站点获取可用信息列表；b）当用户使用浏览器访问搜索引擎，搜索服务器从本地磁盘上的数据库中查找答案

图中，计算机 A 属于搜索引擎公司。它上面运行的蜘蛛程序会访问每一个 Web 服务器（例如，计算机 B），然后将页面中的条目编制成索引，蜘蛛程序会将索引保存在 A 的本地磁盘上。

以后当用户提出搜索请求时，用户所使用的浏览器会作为客户端连接计算机 A 上的搜索服务器，服务器收到用户输入的搜索关键词后，会在本地磁盘上的索引中进行查找，并返回匹配的结果。在这个例子中，计算机 A 上的搜索服务器返回给用户的网页可能存在于任何计算机上（包括计算机 B）。又因为服务器在处理请求过程中不需要与其他计算机通信，只是从本地磁盘抽取信息，因此响应速度会很快。这里面所包含的重要思想是：

因为搜索服务器都是预先收集网页内容信息，并在本地磁盘上形成索引文件。

所以在响应用户搜索请求时，它们不必再去访问 Internet 上的其他计算机。

27.8　现代系统可以搜索页面内容

Internet 上早期的搜索服务都是针对文件名进行查找，用户指定一个文件名，搜索服务返回一系列与文件名匹配的文件列表以便用户下载。现代的自动搜索系统可以针对文件内容进行查找，用户提供一些主题词，搜索服务返回包含这些主题词的网页。目前绝大多数搜索引擎都是针对内容的：

搜索引擎会找出一组内容中含有与用户提供的主题词匹配的网页。

27.9　网页搜索服务的用户界面

对用户来说，网页搜索的使用方法很直观：首先输入一些主题词，经过短暂的等待，搜索引擎就会返回一个与这些主题相关的页面的列表，然后用户可以进一步查看每个页面的具体情况。

用户不需要使用专门的软件进行搜索，查找和浏览过程都可以在浏览器中完成。自动网页搜索服务都有自己的网站，用户只需启动浏览器并访问搜索服务的网站，就可以进行搜索。

⊖　蜘蛛这个名字取得很贴切：当索引程序收集信息时，它会像蜘蛛一样爬遍整个 Web 网。

搜索服务使用的网页通常要包括三部分信息。第一，提供搜索服务公司名称信息和所提供的服务类型。第二，搜索服务的使用说明。第三，用户输入主题词的区域，图 27-2 给出了一个搜索页面的样例[1]。

因为用于搜索服务的网页需要与用户交互，所以这类网页往往会采用表单技术。图 27-2 中的页面就包含了两个特殊的功能区域：一个对话框和一个按钮。使用搜索服务时，用户在对话框中输入主题词，然后点击按钮发送搜索请求。

图 27-2　一个提供搜索服务的网页的样例

27.10　搜索引擎怎样返回结果

前面已经介绍了，自动搜索服务会找出与用户请求相匹配的网页，并创建一个包含与这些页面的链接的新网页。搜索引擎将结果以网页的形式返回给用户浏览器。这类网页不可能采用静态文档，因为不可能预测用户的搜索请求，所以搜索引擎必须动态创建结果网页。图 27-3 给出了一种可能的返回页面。

图 27-3　在用户屏幕上显示搜索结果的示例。这个例子中的网页并非实际存在的

图中说明了当页面内容多于一屏时，浏览器一次会显示的内容。在浏览器窗口的右侧边缘会出现一个滚动条，用户可以通过调节滚动条来决定当前屏幕显示的内容。在这个示例中，搜索引擎返回了 16 个结果项，但是只显示了前 4 项内容。如果页面内容非常多，即使使用滚动条也会很不方便，所以很多搜索引擎每页只显示 20 个结果项，用户可以选择“下一页”来查看更多结果。

27.11　自动搜索使用字符串匹配方法

一些自动搜索服务在寻找匹配网页时采用的是字符串匹配（string matching）算法。搜索服务寻找包含特定字符串的网页。也就是，找到的网页内容必须与用户输入的所有主题词精确地匹配。例如，用户输入的搜索关键词为“automobile”，搜索引擎就会寻找同样包含“automobile”这个词的页面。要点是：

> 当用户在搜索网页的对话框中输入搜索关键词时，他希望搜索引擎能找到包含这些关键词的网页。

27.12　字符串匹配方法的优劣

字符串匹配方法的主要优点在于简单、易用，大多数用户也是希望找到内容匹配的网页。而且，这种方法使用起来很方便。

从开发搜索引擎的程序员的角度看，采用字符串匹配方法省时省力。字符串匹配算法已经被广泛而深入的研究，比较成熟，而且在许多计算机科学的教材中也能找到它的实现方法。更重要的是，研究人员通过一些数学方法可以优化匹配过程，计算机程序能够极快地完成字符串匹配。

字符串匹配方法的主要缺点在于缺少语义方面的考虑：尽管可以精确匹配每一个字，但是

---

[1] 实际中，搜索引擎的页面往往会包含一些广告内容（后面的小节会有介绍）。

字符串匹配程序却不会考虑单个字或词的含义。也就是说，字符串匹配程序只关心字形，不考虑字义。例如，如果用户搜索的是"汽车"(automobile)，那么匹配程序就不会找出包含同义词或相关词语的网页，如"轿车"(car)，"交通工具"(vehicle)等。而且如果用户不小心使用了错别字，很可能就找不到匹配结果。上例中很可能的情况是，自动搜索服务找不到包含"汽车"的网页(由于拼写错误)。

字符串匹配方法的另一个缺点体现在对整句的搜索上。例如，考虑下面这句话：

This sentence does not contain any information about biology money, or foods like butter and milk, and certainly is not about automobile pictures, airline fares, lawyer jokes, opera singers, or library books.

这句话本身没有任何问题，但是字符串匹配程序处理起来却会遇到很大的麻烦，因为它不能理解句子中"not"这个词的所带来的否定含义，因此，可能会误认为该句匹配了对 money、automobile、jokes、opera 和 law 等关键词的查询。

27.13　支持多关键字的自动搜索

尽管字符串匹配可以满足搜索要求，但大多数搜索引擎都提供了更强大的搜索机制。例如，有些搜索引擎允许用户一次输入两个或更多的搜索词语，返回的结果网页可能包含其中的部分或全部的被搜索词。

使用搜索引擎时，用户输入的词语被称为查询关键字(lookup key)，或简称为关键字(key)，因此允许用户输入多个关键字的搜索被称为多关键字搜索服务(multi-key search service)。多关键字搜索的优越性体现在可以搜索一组同义词，例如，用户可能提供了两个关键字：

car automobile

这样可以找出包含 automobile 或者 car 的网页。类似的，可以用多个关键字找出具有同一个主题的页面。例如，对医疗感兴趣的用户可以使用下面的的关键字搜索包含下列任一术语的网页[3]：

health doctor physician hospital medical nurse ambulance

27.14　高级服务可以完成更复杂的搜索

有些自动搜索服务提供的功能超出了多关键字搜索，这些服务允许用户制定更详细的搜索条件。例如，有些搜索服务会找出同时包含用户指定的多个关键字的页面。为了理解这项功能的实际价值，假设有个用户非常喜欢摇滚音乐，所以他很可能使用下面的关键字进行搜索：

rock music

然而，搜索引擎返回的结果很可能会让他失望。搜索引擎会返回包含 rock 或 music 的网页，而不仅仅是有关流行音乐的网页。考虑下面这个句子：

Before you listen to Wagner on a 2500 watt stereo, be sure to warn everyone that the music will be so loud that it will make the entire house rock.

很明显，这句话与摇滚音乐无关，但就因为它同时包含了"rock"和"music"，所以搜索引擎会认为它满足了用户的搜索要求。除了包含上面那句话的页面，搜索引擎返回的结果中还可能有

---

[3] 如果搜索引擎本身不能区分单复数，就可以用多关键字搜索来实现这个功能(同时搜索单数和复数形式)。

关于其他音乐类型的页面，包含关键"Rock Of Ages"的页面，甚至是地理学方面的页面[2]。这里想要强调的是，除了返回用户需求的相关页面外，搜索引擎很可能会返回一些用户根本不需要的网页。

为什么在处理如此明确的查询请求时，搜索引擎的表现会如此糟糕？答案就在于我们所使用的语言。我们的语言经常会产生歧义，主要是我们习惯于用词组表达各种意思（两个词连在一起就构成了词组）。例如，"rock music"放在一起就表示一种音乐类型。虽然搜索引擎也在尝试着去理解词组，但它毕竟还不能很好的处理各种多义词和同义词的情况，因此也就不难理解搜索引擎所犯的各种错误。要点是：

> 尽管搜索引擎在尝试理解自然语言，但它仍然无法识别所有词组，因此，用户必须谨记，在搜索时使用的关键字，除非明确指出，否则搜索引擎会单独处理每一个字。[2]

为了避免歧义的产生，大多数搜索引擎提供了一些高级搜索（advanced search）功能让用户可以更准确的描述所要搜索的关键词。例如，用户可以指定返回的结果页面必须同时包含所有关键词，或至少包含一个关键词，或不包含某个关键词，或匹配整个词组等。而且，有些搜索引擎还允许用户只搜索特定语言（如，英语）书写页，或者它想要查找的文件类型（如，HTML 文件）。这些措施有助于消除歧义。对上面那个有关摇滚音乐的例子来说，用户可以命令搜索引擎将"rock music"作为词组进行搜索，这样返回的页面就一定会包含摇滚音乐的内容。

27.15 个性化的搜索结果

第 26 章介绍的动态网页技术可以提供个性化的网页内容，有趣的是，各大搜索引擎公司都利用了这一点，他们尽可能的收集用户的使用信息，然后利用这些信息控制返回的搜索结果。最终的目的是让搜索引擎能够更好地满足用户的搜索需要。例如，当某个用户搜索"jaguar"时，搜索引擎会判断：如果用户以前搜索过与汽车相关的内容，就会将有关 Jaguar 跑车的网页放在结果中更显著的位置；相对地，如果用户以前搜索过猫、狗之类的动物信息，就会将介绍美洲虎的网页放在更显著的位置。

那么怎样才能够了解用户的使用情况呢？一个简单的方法是在返回的结果上做一些手脚，搜索引擎在返回给用户的结果中并不会提供页面的真正 URL，而是指向搜索引擎上的一个特殊程序。当用户点击这样的链接时，用户的浏览器会首先触发搜索引擎上的一段脚本程序，该脚本会自动记录用户所选择的页面，然后让用户的浏览器自动跳转到真正的页面。对用户来说，这个过程的效果与直接访问页面 URL 相同，搜索引擎只不过充当了一个中间人，但却获得了用户使用搜索结果的情况。下次用户再进行搜索时，搜索引擎就会根据收集到的用户偏好调整返回结果。

总结起来：

> 搜索引擎会保留一些用户信息，并利用该信息选择和调整搜索结果。因此，搜索引擎可以针对不同用户返回个性化搜索结果。也就是说，搜索同样一个关键字，不同的用户可能会收到不同的结果。

27.16 内容搜索的更多细节

前面介绍过搜索引擎的基本工作方式。受 Internet 的规模和增长速度的限制，搜索引擎不可

---

[1] rock 在英语中可以表示"岩石"的意思。——译者注

[2] 这一点主要针对英语，英语中的词组会保留单词之间的空格，在中文不会这样。——译者注

能一次完成对整个万维网的搜索，所以它使用了蜘蛛程序在 Web 上不停收集信息，根据网页内容分门别类地制作索引，然后将索引放在本地磁盘中，来缩短搜索时间。

蜘蛛程序必须检查每个网站上的每一个页面，那它是如何找到这些页面的呢？一种可行的方法是通过记录并跟踪网页中的链接，例如，当蜘蛛程序获取到一个网页时，它会分析 HTML 并提取出所有 URL 信息，进而将这些 URL 加入待搜索的集合中。

Internet 上的网页数量巨大，这意味着搜索引擎公司不可能保存页面中的所有内容。因此，搜索引擎会过滤很多内容，只保留最重要的部分。也就是说，蜘蛛程序在检查一个 Web 页面时，只保留有助于识别该页面的内容，例如，蜘蛛程序会自动去掉冠词和介词，如 a、an、and、the、for 和 with。

27.17　搜索也是有限制的

自动搜索服务使用了另一项重要的优化手段来减少必须存储的信息：只分析页面的前一部分，而不是整个页面，例如，蜘蛛程序可以通过文档的标题获得关键字。因为 HTML 文档的标题保存在特殊的标记之中（<TITLE>和</TITLE>），程序很容易找到它。网页的制作者可以在标题中加入一些关键字，这将有助于搜索引擎找出该网页。一个旅行社的网页可能使用了下面的内容作为标题：

```
<TITLE> Holiday Planner (vacation, travel, Thanksgiving, Valentine) </TITLE>
```

有些搜索引擎还可以识别 META 标记，该标记可以让页面包含一些用户不可见的关键词。例如，旅行社可能会在一个页面中加入下面一些 META 标记：

```
<META NAME= " holiday travel" CONTENT= " travel, airplanes, reservations,
tickets, hotels, rental cars, trains, restaurants, theaters, cruises" >
```

除了标题和 META 标记中的关键字，有些搜索引擎还会从页面的前二三百字中提取一些关键字。所以网页的作者应该尽量将重要内容放在网页的前面。

27.18　广告为搜索服务付账

搜索服务怎样盈利呢？通常用户使用搜索服务是免费的，该服务的主要收入来自广告。一般来说，搜索引擎返回的页面会包含一些与用户查询的关键词相关的广告信息，投放广告的公司也从可这种广告形式。更重要的是，用户使用搜索服务的目的就是为了寻找某些信息，因而很可能对相关的产品和服务产生兴趣。要点是：

> 因为自动搜索服务在每次搜索时都不向用户收取费用，因而在用户看来，网页搜索服务是免费的午餐，实际上其运转是由巨额的广告收入维持的。

不难想到，一些搜索服务公司通过收费方式提高一些网站在搜索结果中的排名。例如，通过与一家汽车制造商签订合同，使该公司的网址第一个出现在所有搜索“汽车”的结果中。总结起来就是：

> 通过付费方式可以提高网页在搜索引擎中的排名（出现在更显眼的位置）。

27.19　自动搜索服务的例子

现在有很多自动搜索服务，而且新的服务还在不断涌现。图 27-4 列出了在本书完成时可用的一些搜索服务。

http://www.100hot.com	http://www.altavista.com
http://www.highback.com	http://www.highint.com
http://www.excite.com	http://www.google.com
http://www.infoseek.com	http://www.lycoscart.com
http://www.lycos.com	http://www.showwww.com
http://www.worldpages.com	http://www.webcrawler.com
http://www.yahoo.com	

图27.4 万维网上可用的自动搜索服务。尽管大部分提供的是通用网页索引，但也有一些只能用来搜索特定类型的信息

27.20 自动网页搜索的重要性

因为Web上的信息正在不断增长，自动搜索引擎在查找信息过程中扮演的角色将越来越重要。通过手工方式查找所需网页的工作量十分巨大，而且找到需要信息的可能性微乎其微。使用搜索引擎，用户可以大大缩小目标范围。

第 28 章　文字、音频和视频通信（IM 和 VoIP）

28.1　引言

前几章介绍了很多 Internet 服务，利用这些服务用户可以发送通知、阅读电子公告板中的消息、浏览 Web 页面、进行页面搜索等等。上述应用，都是用来访问已经存在的信息。本章，我们将要介绍一些允许不同用户之间直接交互的服务。

本章将首先介绍一种能够让两个用户互相发送文本消息的服务，然后会介绍音频和视频通信技术，在解释怎样用软件来录放存储在磁盘上的音频和视频片段之后，还将介绍怎样在 Internet 上传输"现场"产生的声音和图像。最后，介绍一种可以让多人通过 Internet 协作完成某项任务的服务。

28.2　即时消息提供文字交流机会

Internet 上最流行的文字形式的通信服务被称为即时消息（Instant Messaging），简称为 IM。IM 服务让两个用户具有直接通信的能力。任何一方都可以用键盘输入一段文字，然后发送给对方。IM 通常被用来发送一些简短的评论：你发给我一条评论，我再发给你一段回复。

为了更好的理解 IM 服务，假设 Bob 和 Mary 在会议期间使用笔记本电脑通信。图 28-1 画出了 Mary 屏幕上的窗口。

图 28-1　使用即时消息服务时，窗口中可能出现的元素。图中所示为，用户 Mary 正在输入一条等待发送的消息

如图所示，窗口中有两个主要的文字区域：一个用来显示发送过的消息，另一个用来输入等待发送的消息。当用户输入完毕后，点击发送按钮，另一个用户就可以收到该消息。一般来说，会使用不同颜色区分来自不同用户的消息。

28.3　音频和视频功能

音频和视频服务扩展了 IM，使其能够：

- 向其他一个或一组用户发送一段音频
- 向其他一个或一组用户发送一段视频

- 向 Internet 上的所有用户广播一段音频或视频
- 允许一组用户同时查看和修改一篇文档

28.4　音频和视频需要特殊的设备

用户必须拥有容量足够大的 Internet 连接和一些特殊硬件，才能使用音频和视频服务。从技术上说，网络的容量称为带宽（bandwidth），直播视频或音频都需要宽带（broadband）连接才能正常工作，这是因为音频和视频服务产生分组的速度很快。如果网络的容量不足，会使视频或音频变得不流畅，播放过程中出现停顿和跳跃现象。而且，即使压缩之后，视频传输所需的带宽也比音频大 500 倍，因此，可以传输音频的连接可能无法满足视频传输的带宽要求。

除了上面提到的对 Internet 连接带宽的要求，要使用音频和视频服务，计算机还必须拥有：

- 麦克风（microphone），用于捕捉声音。麦克风通常做的很小，可以嵌在计算机的面板内或别在用户的衣服上。
- 耳机或音箱（speaker），用于输出声音。如果需要播放立体声，至少需要两个音箱。音箱可以内置在计算机中，也可以放置在计算机外部。大部分计算机都自带一个便宜的音箱，同时允许外接其他音质更好的音箱。此外，用户还可以选择使用耳机。
- 摄像头（camera），用于捕捉图像。一般会放在显示器上，这样正好可以拍到用户的脸部。便宜的摄像头也被称为 Webcam。
- 显卡（graphics processor），更高速地处理视频和音频内容。

为了理解计算机如何使用这样的硬件，首先要理解计算机是采用电子的方式来控制计算机中的各种设备。例如，用户想要开关摄像头或调节喇叭的声音，他不需要直接操作具体的硬件，只需要使用计算机上的一些软件，敲敲键盘，动动鼠标就可以实现。

28.5　音频剪辑如同一张 CD

计算机中使用的音频是一些保存着对声音进行了数字化处理的结果的文件。这种文件又被称为音频剪辑（audio clip），其内容就是一系列音频采样值。当用户希望播放位于远端计算机上的一段音频剪辑时，需要分两个步骤来完成。首先，要通过 Internet 将整个音频剪辑的副本下载到本机。然后，由本机上的另外一个程序将这一段数字音频采样还原为声音，并通过喇叭播放出来。

28.6　实时意味着没有延迟

计算机专业的科学家和工程师们用实时性（real-time）这个术语来表示一个系统没有延迟。例如，传统的电话和电视系统都是实时的，因为传输在瞬间内完成（用户立即就可用）。但是，获得并播放音频的过程并不是一个实时过程。下载音频的过程本身就是一个巨大的延迟，用户必须耐心等待整个文件下载完成后才能开始播放。

28.7　Internet 上的实时音频

音频可以在 Internet 上实时传输吗？答案是肯定的，但实现起来却很困难。主要原因在于 Internet 所使用的 TCP/IP 协议栈并不是为实时应用而设计的，它并不能保证数据包在一定时间内能够到达目的地。Internet 的工作方式就像是高速公路系统，当多个数据包同时需要发送时，只能有一个可以被立即发送出去，而其他的必须等待。如果没有其他流量，音频数据包会以相同的时间间隔一个接一个的到达。然而，Internet 上有成千上万个可以同时发送数据的计算机，所

以根本不可能在不受干扰的情况下成功发送一个大的音频文件。如果计算机每收到一个音频数据包就播放其中的内容，效果又会是怎样？当长时间没有数据包到达时，我们只能听到一些背景噪声，而当那些被延迟了的数据包和正常数据包同时到达时，我们听到的又是叠加在一起的声音。

令人吃惊的是，Internet 的确可以实时地播放高品质的音频。工程师们给出的解决办法和给大型超市解决供货问题的方法如出一辙。设想有一家大型超市，全天二十四小时营业，其中购买牛奶的顾客络绎不绝。如果我们考察牛奶的卖出情况，可以认为超市中的牛奶是以固定速率在减少。如果我们考察牛奶的进货情况，会发现，每隔一段时间就会有一卡车牛奶被运来。这些新到的牛奶被保存在超市的冷库中，只要超市的冷库存货量足够大，即便运送牛奶的卡车出现晚点，超市也不会出现牛奶脱销的情况。

用来在 Internet 上传输实时音频的软件采用的就是同样的策略。播放实时音频时，接收方的软件在启动后会先接收几秒钟的数据，并将它们放入一个缓冲队列（这个过程与超市将第一次运到的牛奶放入冷库类似）。然后，便开始以固定速率播放声音。播放过程中到达的数据包也会被加入到缓冲队列之中。数据包经常是成批到达，而且每批之间存在延迟。当音频数据包晚点时，软件还能够以正常速度播放声音，这时缓冲队列中的内容会随之减少，但足够维持到下一批数据包到来。图 28.2 表述了这个意思。

图 28.2　播放实时音频的软件所采用的技术。软件在内存中保留了一定量的待播放的声音内容

28.8　Internet 的广播电台

Internet 上的很多服务都采用了上面介绍的音频传输和播放技术。例如，唱片公司以 MP3 数字编码格式发布新专辑中的一些试听单曲，如果用户找到了某首曲子，服务器会传送小样供用户试听。用户还可以找到一些特殊事件的音频，例如音乐会和政治讲演等。

有些服务需要不停地传输实时音频，例如，有些商业广播电台允许用户通过 Internet 收听其节目，还有些电视台也开展了视频和音频的服务。通常，广播电台用 Internet 传送无线电波来播放节目，他们使用一些特殊设备将广播信号转换为数字形式，封装在数据包中，在 Internet 上传输。希望在 Internet 上收听广播的用户需要运行一些软件，这些软件会连接电台的服务器，接收音频数据包，从中提取音频内容，并将其还原为声音信号，最后通过计算机上的喇叭或耳机播放。

28.9　实时音频传输也称为网上广播

为了方便用户收听实时音频，大多数音频都采用了可以被浏览器处理的编码方式。如果进一步希望用浏览器播放实时音频或视频，就必须借助于浏览器插件。这类由软件实现的插件能够完成音频数据包的接收、缓冲和播放。一种称为 RealAudio 的音频技术从众多候选者中脱颖而出，它为大多数浏览器产品都提供了音频播放插件，使浏览器能够接收并播放以 RealAudio 格式编码的音频文件。

还好，安装插件十分简单，浏览器会自动进行下载并安装，不需要用户具备很多计算知识。而且，网络上供下载的大多数插件都是免费的，用户可以在一些网站下载它们，一旦安装了所需

的插件，用户就可以选择一个广播源，并开始收听。这一切和收听普通电台广播没什么两样。

因为大多数实时音频和视频服务都是通过Web浏览器来收听或收看的，所以电台电视台就称这种方式为Web广播（Webcasting）。那用户怎样才能知道节目的时间安排呢？网上有很多提供节目时间表的网站，用户可以访问这些网站。例如，下面的这个URL：

http://realguide.real.com

提供了很多广播源，包括：各个广播电台、电视台、新闻、音乐会以及体育赛事等等。而这个URL：

http://www.audionet.com

提供了另外一些类型的节目列表。

28.10 Internet电话服务会成为现实

很多分析人士认为，Internet上的音频传输技术的进步会将Internet变为一个巨大的电话系统。现在已经有软件可以做到将Internet作为一个私有电话系统来使用。只要两个用户在计算机上都安装了这种软件，他们就可以通过Internet通话。声音在两台计算机之间的传输方式与电话类似。如果两台计算机同时安装摄像头，他们就可以用Internet来传输图像。通过Internet提供电话服务的最大优点在于成本低廉：无论用户打了多少电话，打了多长时间，是否是长途，用户都只需要支付固定数额的网络使用费。

电信公司也意识到了Internet电话的经济优势。各大长途电话运营商纷纷开始在自己网络内部使用Internet的传输方式（IP）。现在很多的长途电话都是使用IP技术。

* 28.11 Internet电话就是所谓的VoIP

Skype和Vonage等公司已经开始面向用户提供IP电话业务，也就是说，不需要传统的电话线，用户只要有Internet连接就可以打电话。拨打电话时，用户可以使用具有麦克风和耳机的计算机，也可以使用一种被称为IP电话（IP telephone）的专用设备（当然需要与Internet相连）。网络界将这种基于IP的电话服务系统称为VoIP（Voice over IP，读作"voyp"）[1]。

当然，如果用户只能在IP电话之间通话，那么这项服务的价值就要大打折扣。为了能让IP电话更具商业竞争力，IP电话的运营公司实现了IP电话系统与传统电话系统的互联互通。这样，每一部IP电话也被分配一个电话号码，用户可以使用IP电话拨打传统电话，也可以用传统电话拨打IP电话。

28.12 音频电话会议

尽管大多数电话都在两个人之间进行，但另一项Internet服务却允许一组用户共同讨论，这类似于电话"会议呼叫"。每次讨论被称为一次音频电话会议（audio teleconference）。为了创建一次音频电话会议，发起者运行用于组织和管理这次讨论的软件，并提供参与者的名单。该软件会自动联系这些参与者。软件或者在每个参与者的屏幕上显示一条消息，或者通知他们有一个电话会议请求，该通知与电话铃音类似。用户可以选择应答请求并加入电话会议，也可以忽略此请求而不参加电话会议。

为了加入电话会议，用户必须运行特定程序来收发音频数据。该程序负责监视用户的麦克

---

[1] 工程师们常用别名"voyp"。

区，将声音信号转换为数字音频编码，并将编码的副本发送给其他参会者。该程序还接收音频数据，还原并播放声音。对每个参会者来说，使用音频电话会议和普通电话会议并没有区别。

音频电话会议的意义在于它能够让其他人感受到说话人情绪。与文字通信不同，人的说话声音可以反映他是喜是怒，是兴奋还是厌烦等等。

28.13　协同文档标记服务

使用声音可以方便大家交流，但通常还需要用文档的形式来确保大家在细节上也能达成一致。比如，假设某个委员会要撰写一份联合报告。开始，委员们可以通过语音方式交换基本观点，但这种方式却不能用来处理细节或生成书面内容。这时，必须要借助文档。

有些软件可以让一组 Internet 用户共同查看和编辑同一个文档。图 28-3 显示了某个用户创建一个共享文档时的屏幕显示情况。

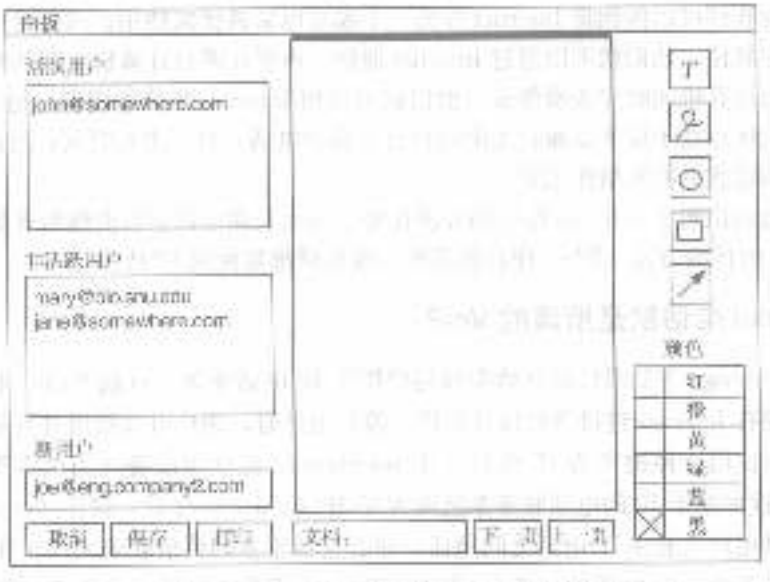


图 28-3　用户创建共享文档时的屏幕输出实例

共享文档服务也被称为白板服务（whiteboard service）。如图 28-3 所示，白板服务软件会在屏幕的左侧列出参与者的名字。图中所示，工作在 somewhere.com 这台计算机上的名叫 John 的用户创建了这次会话，而他正在邀请参会者。为了添加新的参会者，John 需要在名为“新用户”的空白区域中输入他们的登录名和计算机名。系统会向每一位用户发送邀请，同时会将他们的信息移至“非活跃用户”列表中，并等待他们的应答。当一个新用户加入当前会话，系统会将他的名字自动移至“活跃用户”列表中。

要显示文档，需要有一个用户在标记为“文档”的输入框中填入本地磁盘上一个文件的名称，然后白板服务会在每个参会者的屏幕上显示该文档的内容。图 28-4 显示了当参会者多于两人而且指定了等待编辑的文档后的情况。Joe 出现在“非活跃用户”列表中，这说明有人邀请他加入当前会话，但是他还没有给出答复。任何参会者可以随时退出会话；当某个用户退出后，白板服务会将他的名字移至“非活跃用户”列表中。

28.14　对文档进行标记

用户使用鼠标与白板服务进行交换。例如，任何用户可以通过点击“下一页”或“上一页”命令按钮来实现翻页。而且，用户可以在任何时刻向文档添加绘图或文字。当一个用户翻页或修改文档后，所有用户都会在自己的屏幕上看到这种改变。实际上，

> 白板服务可以看作为用户提供了一张白纸，所有参与者都可以看到纸上的内容，而且可以立即看到任何人对它所做的修改。

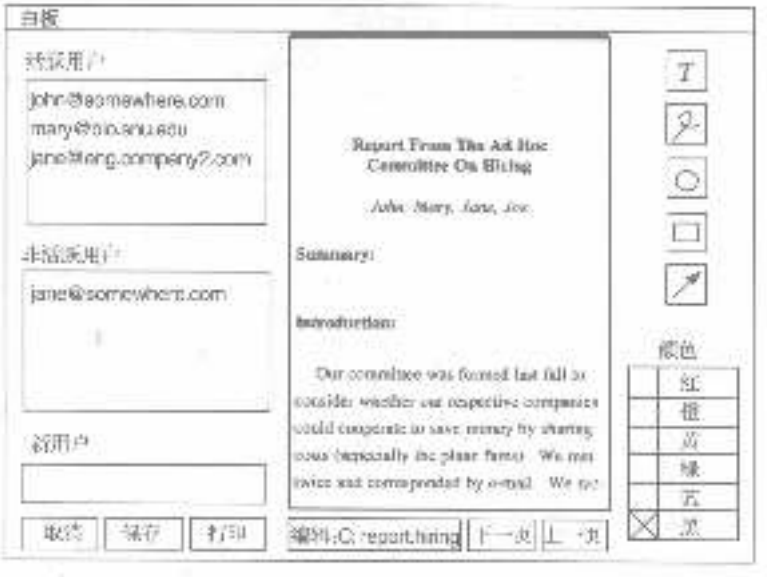


图 28-4　白板服务打开了某个文档，所有参会者都会看到同样的输入

图 28-4 说明了用户怎样控制一个白板会话。屏幕的右侧包含了设定绘图模式的方框。用户可以选择带有“T”标记的方框来输入文字；选择第二个方框可以画任意曲线；选择第三个方框可以画圆；选择第四个方框可以画矩形；选择最后一个方框可以画箭头。

除了可以选择绘图模式，用户还可以通过屏幕右下角的列表选择颜色，以后该用户对文档的修改都会用这种颜色来显示。

28.15　参与者共同讨论并标记文档

通常，白板会话的参与者们同时也在使用音频电话会议。这样，他们就可以一边讨论一边看屏幕上的文档。例如，正在共同讨论图 28-4 中的文档的人们决定，将其中委员会成员的名字改用全名，而且讨论过程中，Mary 同意由她完成总结部分，另外有人发现“especially”的拼写错误。任何参会者都可以在讨论过程中修改文档。也许是 Jane 发现了文档中的拼写错误，并用圆圈打上标记。John 增加了一段注释来提醒大家修改人名的决定。随着讨论的进行，大家不断修改文档，最终成为如图 28-5 所示的那样。但是大家所做的修改并不会自动保存在原始文档中，它们仅存在于白板上。所以，最后还需要有人用字处理软件修改原始文档。

使用彩色显示器的用户还可以用不同颜色来强调他们画的图和作出的评论。例如，在白底黑字的文档中，红色标记会十分醒目。而且，如果不同人采用不同颜色进行修改，就可以通过颜色来锁定评论的作者。另外，大家还可以规定用某种颜色来表示问题和建议，而用另一种颜色表

示最终决定。例如，图 28-5 中，标记拼写错误的圆圈可以用红色，来表示该修改是必需的，而“USE FULL NAMES”的注释在大家达成一致之前可以用蓝色。

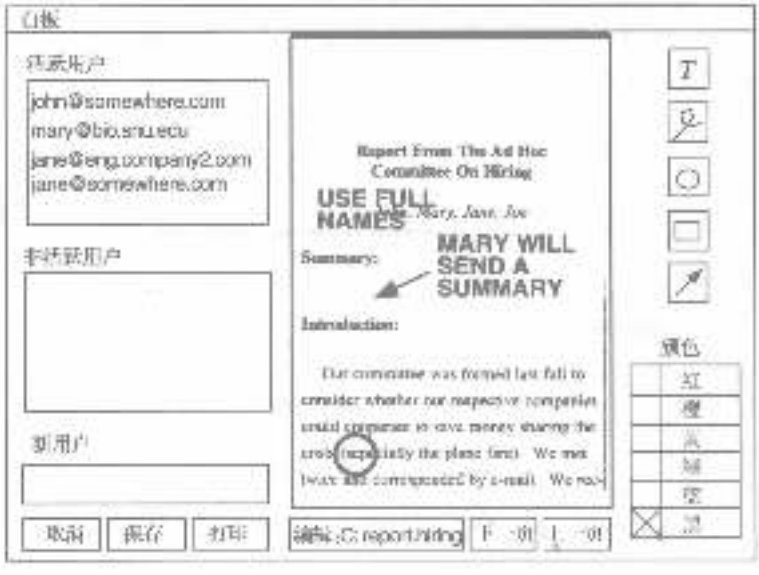


图 28-5　本例展示了参会者怎样在白板上修改文档，白板服务会将当前会话的结果保存到磁盘上，便于日后重新查看

28.16　视频电话会议

音频电话会议和共享白板服务可以支持多人协同工作，这在准备或修改文档时特别有用。但是，当需要一组人集思广益时，面对面的交流方式会取得更好的效果，因为表情可以节省很多不必要的沟通。

为了实现面对面的交流，Internet 提供了视频电话会议（video teleconferencing）服务。创建视频电话会议的过程和白板服务类似：一个用户首先发起会话，并邀请其他用户加入，最后把所有人联系在一起。当新用户加入时，他的视频图像就会出现在每个用户的计算机屏幕上的一个小窗口中，显示效果和普通电视一样。当一个用户微笑或皱眉，其他用户都可以看到[1]。

当然，只有图像还不能让用户进行有效交流，因此，大多数视频电话会议系统都会同时提供语音和视频信号。当一个用户加入会议后，他可以同时看并且听到其他用户的情况。

28.17　小组视频电话会议

视频电话会议在用户较少时效果很好，每个用户都可以通过屏幕看到其他用户的图像。但是，当画面比较多时，就难看到所有用户的头像。

当很多人一起参加视频电话会议时，屏幕不足以显示所有用户的图像。这时，所有参与者必须划分为几个小组，每个小组占据一个房间，每个房间配备一台摄像头和一个更大的屏幕。实际上，每个房间是电话会议的一个成员单位，摄像头必须捕捉该房间图像，然后发送给其他用户

[1] 价格低廉的摄像头的质量当然不如专业的电视设备。如果用户的网络连接速度很低，图像还可能会发生抖动或跳跃等情况。

组，每个小组的显示屏也是分别显示其他组的图像。

同时捕获整个小组的声音比较困难，单个麦克风不能侦测到房间中的所有声音，而采用多个麦克风时，又会出现比较大的背景噪声，为了解决这个问题，专门用于视频会议的房间一般都配备了可移动的麦克风，这样可以方便地将麦克风传到发言人的面前。

28.18 同时提供音频、视频、白板和即时消息服务

当联合使用音频和视频电话会议、白板和即时通信服务时，就构成了一种更加吸引人的服务。为了了解它们使用效果，假设每个用户的显示器都和普通电视一样大，屏幕的四周的矩形小方框显示每个远程用户的图像信息（通常只包括面部）。屏幕的中央区域用户白板服务。最后，用户还可以使用及时通信软件私下交谈。

视频、音频、白板、即时消息的联合使用，不仅可以让所有用户都能看到文档内容，而且还可以相互看到彼此的头像并听到他人的声音，当用户需要讨论时，他们可以使用视频和音频通信；当用户需要确认对文档的修改时，他们可以使用白板服务；当用户之间需要私下沟通时，他们可以使用 IM。事实上：

> 实践证明，联合使用白板、视频和 IM，可以获得最大限度的灵活性，因此很多电话会议系统同时提供这三种服务。

28.19 小结

Internet 提供了多种可以允许两人或多人交互的服务，即时消息可以让一对用户交互文本消息。

音频会议服务与多方电话呼叫类似，两个或两个以上用户可以建立电话会议，并相互交谈。使用者还可以让其他用户加入讨论。视频电话会议与音频电话会议也类似，只不过将通信的内容从音频变为了视频，采集设备也换成了摄像头而已。为了显示其他用户的视频图像，计算机将屏幕分割为多个矩形区域，每个区域显示一个用户的图像。当参加视频会议的人数很多时，他们分散到不同的小房间中，每个房间都配备摄像头和麦克风，然后以房间为单位加入视频会议。

白板服务允许一组用户一起评论和标记同一个文档，某个用户创建白板会话并指定一系列的参会者，每个参会者会收到一份邀请，而且可以随时退出。白板服务还为每个活跃的参与者提供一份共享文档的拷贝，并允许用户使用各种颜色修改文档，所有用户都可以看到修改的结果。

28.20 个人见解

尽管视频会议服务十分令人期待，但它会消耗比其他服务更多的网络资源。视频和音频传输与普通的数据传输服务有很大的不同，在 Internet 过载的情况下，普通数据传输可以减慢或暂停，而视频和音频传输却不能，否则图像和声音会中断。因此，电话会议会造成低速 Internet 连接过载。

当我在写这本书的第 1 版时，曾经使用过音频电话会议和白板服务，会议的参加者之一在某所大学内使用了一条低速 Internet 连接，我们的这位参会者同事时不时地就会从对话中断开。最后我的一个朋友——一个计算机专家，通过测量网络流量找出了问题的根源：我们的电话会议使得他所在大学的 Internet 出口完全被占满，后来再次使用电话会议时，我们想了一些减少流量的方法。但最终我们无可奈何的接受这样的事实：如果想要保持连接不断开，我们就必须尽量少说话。

今天，很多大学的 Internet 出口带宽都很大，足够支持音频电话会议，但对于普通用户，使用视频会议的效果还很差，经常会出现抖动、停顿等现象。即使经过压缩，视频产生的流量仍然很大。这种情况下，必须进一步减少数据的传输数量，但这是以牺牲图像质量为代价的。

第 29 章　传真、文件传输和文件共享（FTP）

29.1　引言

前几章介绍了一些用户可能经常用到的 Internet 服务。本章我们再来介绍三种服务。尽管这三种服务并不像前面介绍的服务那样家喻户晓，但是从另一个方面来看，它们展示了 Internet 上服务的通用性和多样性。

29.2　发送传真

尽管传真（fax）技术从发明到现在已经几十年了，但直到传真的发送速度因现代通信技术的进步而获得大幅提升之后，它在商业活动中才获得了广泛应用。到 20 世纪 90 年代末，大部分公司以及有许多个人用户都拥有传真机，传真机传送服务得到了广泛应用。

传真是一种简单的数字传输方式，它使用常见的声音电话系统进行传输。传真机主要由 4 部分构成：打印模块、扫描模块、起拨号作用的调制解调器（Modem）和起到控制作用的一个专用计算机。调制解调器用于收发传真。发送传真时，首先用调制解调器拨对方的号码，等待对方的传真机应答。当双方调制解调器完成同步后，发送方用扫描模块将待传内容转化为电子文档的形式，并利用电话线传输到对方传真机上。然后对方传真机再利用打印模块将所需内容重新打印成文档。

在 Internet 的早期，有人开发了特殊程序让计算机能够通过 Internet 发送传真。尽管目前传真服务仍然存在，但是已经很少使用了，取而代之的是利用电子邮件传输文档。总结起来：

> 尽管在 Internet 上可以发送传真，但越来越多的人选择电子邮件的方式来发送文档。

29.3　可以通过 Internet 拷贝文件

尽管可以通过电子邮件或者使用即时消息软件发送一些不太大的文档，但是它们都不适合发送大量数据。Internet 上有一种专门的服务能够高效地将任意大小的文件从一台计算机的磁盘上传输到另一台计算机的磁盘上。用户可以用这种服务下载（即获得文件副本）包含音乐或视频文件。下面几节将详细介绍文件传输服务软件的工作原理。

29.4　文件中的数据

计算机将大量的数据保存在存储设备中，例如磁盘。从用户的角度来看，磁盘上的数据是以文件夹（folder）和目录（directory）的形式来管理的。大多数系统中，文件的大小是可变的：一个文件可能很大也可能很小，还可以随着用户添加或删除文件的内容而增大或减小。例如，文字处理软件通常将每个文档保存为一个单独的文件，一个文件可能只有几行，也可能包含上百页内容。

29.5　拷贝文件

使用 ARPANET 的研究人员很快就发现，利用网络可以在不同计算机的磁盘之间拷贝文件，

因此他们开发了可以实现这一功能的软件，并命名为文件传输（file transfer）。不久之后，他们又重写了该软件，使其可以在Internet上工作，这就是我们今天经常使用的文件传输协议（File Transfer Protocol），简称为FTP。

29.6 在浏览器中使用FTP

起初，要使用FTP服务，必须运行特定的程序。现在，几乎所有的浏览器都集成了这一功能，用户只需要点击网页上的文件链接就可以将其拷贝到自己的机器上。浏览器通过判断链接的URL来决定是否使用FTP。如果URL以ftp://开头，浏览器就会知道这是一个文件传输的请求。例如，下面的URL指向一个可以通过FTP获得的文件，该文件存在于Purdue大学的一台计算机上：

ftp://ftp.cs.purdue.edu/pub/comer/tib/example.txt

当用户输入上面这个URL，浏览器会联系名为ftp.cs.purdue.edu的计算机上的FTP服务器，并请求一份pub/comer/tib/example.txt文件的拷贝，然后将文件内容显示给用户。

总结起来：

> 文件传输协议可以用来下载大文件，尽管有很多单独的软件可供使用，但是大多数用户是直接通过浏览器来访问FTP，浏览器根据URL的前缀（"ftp://"）来判断是否使用FTP协议。

29.7 通过FTP浏览目录内容

除了获取特定文件之外，FTP还允许用户浏览目录内容（查看文件夹），当用户输入的URL指向的是目录时，浏览器会请求FTP服务器返回该目录所包含的文件列表，然后显示给用户。FTP目录列表的显示方式有别于普通的Web页面：为了美观，Web页面会采用悦目的布局和绚丽的图片，而FTP目录只是简单的包含一个标题和文件名列表，每个文件名都对应下载该文件的链接。因此，FTP站点中文件的名称必须做到一目了然，便于让用户了解文件的内容。

通过实际例子可以让读者更好的理解这些概念。为此，我特地建立了一个小型的FTP目录，读者可以通过下面的URL来访问：

ftp://ftp.cs.purdue.edu/pub/comer/tib

这个目录包含了三个文件：

Crazy_Example.ps
README
example.txt

通过访问这个实际的FTP站点可以看出：

> 尽管用户可以通过FTP浏览目录中的内容，但是能够看到的只是文件名列表而已。

29.8 通过FTP上传文件

前面只是介绍了用户如何通过FTP获得远程机器上的一份文件拷贝。的确，这也是绝大多数用户使用FTP的目的。我们用下载（download）来表示这个过程，也就是说，用户可以下载文件。

有意思的是，FTP还可以让用户进行反方向的文件传输：将本地的一个文件拷贝到远程计算机上。我们用上传（upload）来表示这个反方向拷贝过程。当然，用户要利用FTP上传文件，

就必须满足一些条件。首先，远程机器要开通 FTP 上传服务，其次，用户必须拥有相应的权限。

29.9　通过 FTP 传输文件必须经过授权

那么如何才能只允许一部分用户通过 FTP 访问文件呢？例如，收费音乐下载网站怎样防止没有付费的用户下载文件？FTP 是通过引入权限检查来解决上述问题：任何人想要访问 FTP 服务器，都必须提供合法的用户名和密码，否则就不能进行下载。而且，FTP 站点的管理员还可以为不同用户分配不同的文件访问权限。因此，可能一个用户可以下载 A 文件，却不能下载 B 文件；或者只能上传文件，而不能下载任何文件。

如果每个 FTP 会话都需要用户提供用户名和密码，那么在无须提示用户输入登录信息的情况下，浏览器又是怎样做到直接下载文件的呢？原来 FTP 允许用户匿名登录（anonymous login）来访问一些可以公开的（public）文件。一般来说，FTP 服务器都允许用户使用一对公开的用户名"anonymous"和密码"guest"访问公开文件。因此，当用户点击了指向 FTP 文件的链接时，浏览器会自动尝试匿名登录，如果失败，就会提示用户输入用户名和密码。当然，匿名登录的用户只能访问那些已经公开的文件，而无权访问 FTP 服务器上的其他文件。总结起来：

> 远程 FTP 服务要求用户必须提供正确的用户名和密码，才能获得访问 FTP 资源的授权。如果只需要访问公开文件，可以采用"anonymous"作为用户名，"guest"作为密码。

29.10　FTP 工作原理

与 Internet 上的很多服务一样，FTP 也采用了客户端—服务器（client server）的工作模式。当用户输入表示 FTP 服务的 URL 时，用户的浏览器就作为 FTP 客户端与远程计算机上的 FTP 服务器建立 TCP 连接，每次用户请求文件时，FTP 客户端和服务器配合完成这个下载任务，通过 Internet 将文件副本传输到客户端。图 29-1 描述了 FTP 下载过程中数据的流动情况。

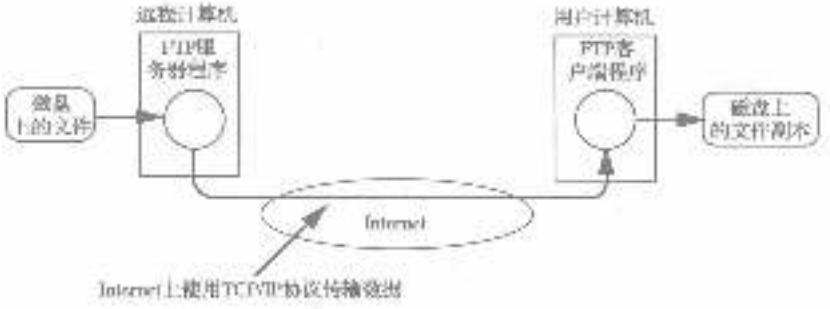


图 29-1　服务器响应用户请求，给客户端发送文件副本过程中数据所经过的路径。FTP 客户端和服务器使用 TCP/IP 协议进行通信

首先，FTP 服务器在本地查找用户请求的文件，通过 TCP 连接将文件内容传送给客户端。客户端收到数据后，写入本机磁盘上的文件副本中。当全部内容传输完毕后，客户端和服务器关闭 TCP 连接。客户端（例如，用户的浏览器）可能会调用其他程序显示文件内容。

29.11　FTP 的影响和重要性

在 Internet 的初期，经过 FTP 传输的数据在 Internet 的总流量中占据了相当大的比重。实际上，

早期 Internet 上的 FTP 流量比其他任何一种服务产生的流量都多。直到 1995 年，万维网服务才取代 FTP 成为 Internet 上产生流量最多的服务。

29.12 对等方式的文件共享

从 2002 年开始，Internet 上兴起了一种新的文件传输方式，其名曰：文件共享（file sharing）。这个新应用的灵感来源于用户之间可以交换各自的音乐文件。如果有两个用户 A 和 B，其中 A 有一张专辑，而 B 有另一张专辑，那么 A 和 B 只需要简单的相互拷贝对方的音乐文件就可以同时拥有两张专辑。但是很快，这种最初被设计用来在朋友之间分享音乐的应用，便被更多的人用来（非法）交换各种受版权保护的文件。不久，这种新的应用就完全被用来在相互陌生的用户之间进行文件共享，共享的内容主要是数字音乐和视频文件。

这种文件共享服务是怎样工作的呢？与传统的 Internet 服务不同，文件共享服务无须使用功能强大但十分昂贵的服务器来处理用户请求，小型、慢速的计算机就可以胜任。加入文件共享服务意味着用户同意让自己的计算机承担双重设备的角色：为了能够下载自己所需的文件，用户必须为其他用户提供上传。这样一来，用户不必都从同一台服务器下载文件，而是可以从其他用户的计算机下载。文件共享软件会记录本机上共享的文件，以及查找其他文件下载地址所需的必要信息。当一个用户收到对某文件的请求时，他首先会检查本机是否有该文件，如果有，就给对方上传，如果没有，则返回一些可能包含所请求文件的计算机的信息。

前面介绍过，Internet 本身就被设计为一种对等网络，即一台计算机可以和其他任意一台计算机通信。而文件共享传输正是利用了这一特点，在任意两台计算机之间共享文件，所以这类应用也被称为对等应用（peer-to-peer application），简称 P2P。对等通信模式的功能十分强大，利用这种模式，可以很容易地在 Internet 上推出新的服务。然而，通过文件共享的例子可以看出，这种能力也可以被用来支持非法活动。总结起来：

> 文件共享是一种对等应用，因为文件传输是在任意两个计算机之间进行的。
> 但共享的文件常被用以非法传输受版权保护的内容。

29.13 小结

可以利用 Internet 将文件从一台计算机拷贝到另一台计算机。FTP 就是一种专门用来在 Internet 上传输大文件的服务。大多数用户使用浏览器来访问 FTP 服务器，当浏览器遇到以 ftp://开头的 URL 时，会自动充当 FTP 客户端的角色。

尽管大部分用户只是使用 FTP 下载文件，但 FTP 还允许用户将文件上传到远程计算机上。而且，通过 FTP 服务，浏览器还可以让用户查看远程计算机中某个目录下的文件列表，并允许用户选择下载某些文件。

因为文件共享可能发生在任意两台计算机之间，所以它是一种对等应用。加入共享服务的计算机在下载其他计算机上的文件的同时，必须提供一些文件共享给其他用户。但是文件共享往往被用来传播盗版内容。

第 30 章　远程登录和远程桌面（TELNET）

30.1　引言

前几章已经介绍了很多 Internet 服务，本章将介绍一种可以让用户访问远程计算机的服务。除了讨论该服务的工作原理外，还会介绍它的意义所在。

30.2　早期计算机使用字符界面

要理解 Internet 上基本的远程访问服务，首先要了解计算机的字符界面。在 Internet 刚刚诞生的时代，计算机主要是那些分时（timesharing）计算机。与现在每个用户独享的个人电脑（PC）不同，一个大型的分时系统允许很多人用时使用一台计算机。每个用户通过各自的终端（terminal）来使用分时计算机。一般来说，一个分时计算机会和很多个终端相连，而每个终端只包含键盘和显示器。

从用户使用的角度来看，分时计算机和 PC 并没有差别。分时计算机上运行的特殊软件让每个用户都觉得自己在独享计算机。用户可以决定什么时间运行什么程序。这些计算机的处理能力一般都很强大，可以同时处理多个计算任务。例如，同一时刻允许不同用户分别运行电子表格和字处理程序，而且，分时系统还会为每个用户单独分配磁盘空间，以便保存文件。

与 PC 机一样，分时系统会立即响应用户的输入。例如，当用户敲击键盘或移动鼠标后，屏幕上会立即反映出这些变化。实际上，分时系统工作得非常出色，用户根本无法察觉是否有其他用户在使用同一台计算机。除非两个用户同时要使用一个资源（例如两个用户同时使用打印机，必然有一人要等待另一个打印完成后才能打印）。

30.3　分时系统需要为每个用户分配标识

因为是多个用户同时使用分时系统，所以使用前每个用户必须拥有一个独一无二的身份标识符。实际系统使用的是为每个用户分配的登录用户名（login identifier）和密码（password）。当用户想要使用计算机时，必须首先输入用户名和密码。这个过程就称为登录（logging in，log-in）。

用户成功登录到分时系统上之后，就可以使用命令来执行需要的程序。当用户完成任务，不再需要使用计算机时，通过注销（logging out）命令通知系统。

总结起来：

> 分时系统允许多个用户同时运行程序。每个用户拥有自己的用户名和密码。
> 如果用户需要使用计算机，必须首先正确输入用户名和密码。

30.4　远程登录和本地登录几乎没有差别

用户可以利用 Internet 上的远程登录服务来远程使用分时计算机系统，其效果就相当于将本地键盘和显示器与远程计算机直接相连。使用远程登录服务时，用户需要在本地计算机上运行一个特殊程序并指定一台远程计算机。该程序会创建一个窗口，并与远程计算机建立连接。此

时，用户在该窗口中所进行的任何键盘输入操作都将被传送到远程计算机上进行处理，任何处理结果也都将被传送回本机，并在同一个窗口中进行显示。实际上，这个特殊的窗口就好像是直接连接在远程计算机上的一个终端。

30.5 远程登录工作原理

与其他 Internet 服务一样，远程登录也是采用客户端—服务器模式。当本地计算机上的用户想要登录一台远程计算机时，他只需运行一个程序。这个程序作为客户端和位于远程计算机上的服务器建立一个基于 TCP/IP 的连接。服务器会将这个网络连接看作一个普通的本地终端（即，服务器首先提示用户输入用户名和密码）。图 30-1 解释了这个过程。

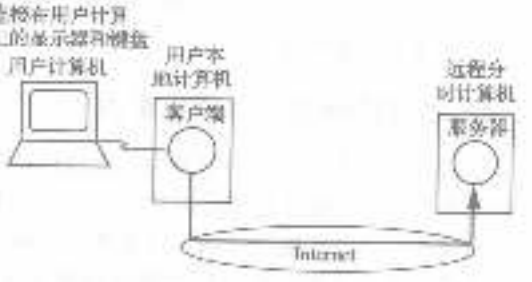


图 30-1 远程登录服务涉及两个程序。用户运行本机上的一个程序作为远程计算机的一个客户，键盘输入传输到远程计算机上去处理，而传回的输出结果则显示在用户屏幕上的一个窗口中

30.6 远程登录过程中的转义问题

尽管远程登录的客户端程序通常将用户的每一个键盘操作都发送给远程计算机进行处理，但它还必须提供一种暂停发送的方法，以使能够在本地计算机上处理键盘操作的方法。要理解这种机制存在的必要性，必须深刻认识下面两个事实：

- 在远程登录期间，至少有两个程序在同时运行：正在远程计算机上运行的程序和本地计算机上的远程登录客户端程序。
- 大多数操作系统都保留了一些具有特殊功能的组合键（例如，Windows 系统中的 Ctrl+Alt+Del）。

问题是：当用户按下某个组合键时，应该将它交给远程计算机来处理，还是应该由本机来响应？实际上，用户需要同时控制本地和远程计算机。因此，远程登录服务保留了一个特殊的转义序列[1]。转义序列之后的键盘操作将不会发送给远程计算机，而是由本机进行处理。其他情况下的键盘操作都可以看作是对远程计算机的输入。转义序列的含义就是告诉本机上的远程登录客户端程序："请先暂停对远程计算机的输入，后面的操作请用于控制本地计算机。"

30.7 Internet 远程登录的标准是 TELNET

Internet 上远程登录服务的标准是 TELNET 协议，它是 TCP/IP 规范的一部分。TELNET 协议详细的规定了远程登录客户端与服务器之间的交互过程。例如，客户端如何与服务器建立连接，客户端如何将击键动作编码传输，而服务器又如何对输出结果进行编码传输。

因为 TELNET 客户端和服务程序使用相同的规范，他们在通信细节上是一致的。例如，尽管大多数计算机都在键盘上设定了一个特殊的终止（abort）键，使用它可以终止当前正在运行的程序。但并不是所有计算机系统的终止键都使用相同的键，有些计算机使用的是 ATTN 键，有些则是 DEL 键。TELNET 用一段特定的比特序列来代表终止键，当用户敲击了本地键盘上的终

[1] 尽管键盘上的 ESC 键最初是被设计用来完成转义功能的，但在很多程序中 ESC 键有其他的含义，所以 Internet 上的远程登录程序使用 CONTROL-] 来表示转义。

止键，TELNET 客户端程序会自动将其编码为特定比特序列。要点：

> 因为 TELNET 制定了传输标准，所以它允许用户在使用不同键盘的计算机之间进行远程登录。

30.8　远程访问也可以提供桌面

早期计算机大多使用字符界面。而现代计算机可以在屏幕上显示多个窗口、图形图像、图标和下拉菜单。很多程序也都支持使用鼠标进行操作。因为 TELNET 是针对字符界面而设计的远程访问服务，所以不能用来远程显示图形界面。

幸好，现有技术已经可以处理图形界面元素。利用这种技术可以实现远程桌面（remote desktop）。远程桌面系统可以获得一台计算机上的屏幕显示情况，并将副本发送给远端的另一台计算机。两台计算机上都需要运行特定的软件，其中一个负责拷贝当前屏幕的输出情况，另一个则负责将收到的内容拷贝绘制到本地屏幕上。这样一来，用户在本地计算机的屏幕上就可以看到与远程计算机相同的图像。图 30-2 解释了这个概念。

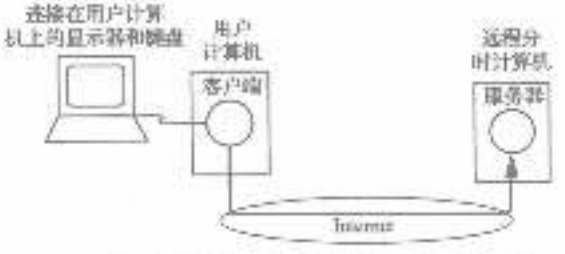


图 30-2　远程桌面服务，通过软件可以在用户本机的屏幕上显示远程计算机上的图像

30.9　远程桌面的运行机制

远程桌面服务的运行机制和远程登录服务类似。要使用远程桌面，用户首先运行客户端程序，输入远程计算机的地址，等待与远程计算机上的服务器建立连接。程序会接管用户本机的屏幕、键盘以及鼠标[6]。这时无论用户移动鼠标还是敲击键盘，都将被发送到服务器进行处理。

服务器运行在远程计算机上。它负责获取远程计算机的屏幕输出信息，并传回给客户端，以便用户能够看到同样的屏幕图像。与此同时，服务器还要接收客户端发送的键盘鼠标输入，并把它们当作本地输入一样进行处理。

远程桌面服务最吸引人的特点莫过于身临其境般的使用体验。通常，当计算机启动后都会显示桌面，桌面上有很多代表各种含义的图标，例如“我的电脑”。通过远程桌面服务，用户可以自由查看，随意控制远程计算机的整个桌面。例如，如果用户决定重启计算机，那么远程计算机就会重新启动；如果用户将一个文件拖放到回收站，那么位于远程计算机上的该文件就会被删除。

有些远程桌面产品可以免费使用，另一些则必须付费。例如，RealVNC 公司发布了免费的 VNC（Virtual Network Computing）软件以及功能更强大的收费版本。要获得更详细的信息可以访问该公司的主页：

http://www.vnc.com

30.10　远程登录和远程桌面的重要性

Internet 上的远程访问服务的重要性主要体现在三个方面。第一，远程访问使得远程计算成

6　有些远程桌面软件只能在一个小窗口中显示远程桌面，这样文字和图片都会相应变小，而且需要用户使用滚动条来查看窗口的不同部分。

为现实。与其他只能传送消息或文件的服务不同，远程访问允许一台计算机上运行的程序接收并处理来自其他计算机的输入，然后将输出结果反馈给远端用户。

第二，远程访问具有通用性的特点。一旦用户远程登录到其他计算机，就可以执行该计算机上的任何应用程序。这样一来，无论用户使用什么型号的计算机，都可以通过远程访问来运行那些只能在特定品牌计算机上运行的程序。

第三，远程访问具有大量的用户。现在，很多公司都在大量地使用各种计算机软硬件设备，其中很多程序和数据库都只能从所在计算机上运行或访问。如果不使用远程访问，工作人员就必须不停的奔转于各台计算机，而使用远程访问后，他们只需坐在一台联网的计算机前，就可以远程完成各项工作。

30.11　通用性让远程登录和远程桌面更强大

一旦用户成功登录远程计算机，便可以运行其上的任何应用程序。尽管程序实际运行在远程计算机之上，但是用户却丝毫感觉不到，完全相同的输入和输出效果让用户感觉自己就坐在远程计算机之前。

远程登录服务的另一个强大之处是：在不修改现有程序的情况下，让用户能够运行任何一台计算机上的任何一个应用程序。用户只需要额外运行一个远程登录程序，便可以使用其他计算机上的程序资源。下面通过一个实际例子来说明这一点。

假设有一个公司使用数据库来保存它的产品信息、价格以及库存情况。并且该数据库运行在计算机 X 上，公司的销售人员使用该数据库为客户提供价格和货物信息。

假如公司又希望让潜在客户能通过 Internet 访问该数据库。但是很可能，数据库软件既不支持 Internet，也不允许用户在其他计算机上使用。更重要的是，公司已经花重金用于购买该数据库软件，培训员工使用该系统，因此不可能更换功能更强大的数据库（如，支持 Internet 访问的数据库）。

远程访问可以帮助该公司解决燃眉之急。他们只需要在计算机 X 上安装一个远程登录服务器程序，并添加一些远程用户账号。例如，公司可以创建一个可以访问数据库的公开账号，也可以为每一位客户分配一个单独的账号。

30.12　远程访问可用于多种类型计算机

Internet 提供的远程访问服务解决了一个重要问题：使得不同品牌的计算机相互通信成为可能。例如，某个公司的数据库软件只能在 IBM 的计算机上工作，但可能销售人员必须使用其他品牌的计算机（如，苹果电脑）来访问公司的数据库。在这种情况下，公司员工就可以利用远程访问来操作数据库。总结：

> 远程访问服务允许一台计算机上的用户使用另一台计算机上的应用程序。该服务的强大之处还在于：不需要对已有的应用程序做任何修改。

30.13　远程访问的一些异常结果

尽管远程访问带来了很多方便，但有时访问结果也会给用户造成困惑。问题源于用户坐在一台计算机前面运行的却是另一台计算机上的程序。当远程桌面启动时，用户无法察觉，用户无法仅从屏幕的输出区分所看到的是远程程序运行的结果还是本地程序运行的结果。

我们通过一个例子来说明远程桌面可能带来的麻烦。考虑有一个用户登录到远程桌面，并运行字处理软件。创建了一个文档之后，用户希望将它保存在磁盘上。虽然看起来和本机上运行

的字处理软件完全一样，但是实际上，当前的程序是运行在远程计算机之上的。因此用户可能会错误地认为：既然我可以在自己的屏幕上查看文档，我就可以在与本机相连的打印机上打印它。但是，与用户"看"到的情况不同，字处理软件是运行在远程计算机之上，因此只能使用与远程计算机相连的打印机。与此类似，当用户命令远程字处理软件保存文档时，文档是被保存在远程计算机的磁盘上。因此，尽管远程桌面中的内容看起来与本机中的内容没什么不同，但用户却不能使用本机上的资源（打印机等）对它们进行处理。

要点是：

> 远程登录用户需要谨记：尽管可以使用本机的显示器、键盘、鼠标来控制输入输出，但应用程序却只能操作远程计算机上的文件和打印机。

30.14　小结

远程访问机制允许用户在一台计算机上运行另一台计算机上的程序。Internet 上的第一个远程访问标准：TELNET，仅提供了字符界面方式的远程登录。更新的技术实现了图形化的远程桌面，让用户感觉自己就好像是在远程计算机上进行工作一样。

对于需要同时访问很多台计算机的用户来说，远程访问是一棵救命稻草。他们不必再奔波于各种计算机之间，只需远程登录每一台需要的计算机即可。在进行远程访问时，本机的键盘、鼠标和显示器就好像直接与远程计算机相连。与用户"看"到的不同，远程访问过程中使用的文件和打印机也都是位于远程计算机之上，而非本地计算机。

第31章 安全通信技术

31.1 引言

前几章我们介绍了一系列Internet服务，本章及后面两章主要讨论怎样在实际中安全使用这些服务。本章首先引出一个基本概念：需要一些手段来保证Internet上通信的安全，然后讨论安全性的含义和需要它的原因，接下来介绍一些保证通信安全的关键技术。第32章会继续就安全话题进行讨论，并介绍一种可以保证出差或在家中工作的员工安全使用网络的技术，以及怎样保证电子商务的安全性。

31.2 Internet并不安全

组成Internet的网络绝大部分都属于"共享"网络，也就是多台计算机连接在一起的网络，共享网络的主要缺陷在于缺乏安全性：网络上的任意一台计算机都可以窃听其他计算机的通信内容，因此，这种网络不安全（unsecure）正是因为组成Internet的网络本身大多并不安全，所以总的来说Internet也是不安全。

第14章介绍了一些不安全的Internet接入技术：有线电缆解调器（Cable modem）和无线网络（Wi-Fi）。拿有线电缆解调器来说，每一个用户处于由上百个用户构成的组中，与组内的其他用户共享网络带宽，正因为如此，使得在线电缆解调器中窃听其他用户通信成为可能：用户只需要装一个很便宜的软件就可以收到组内其他成员的所有通信数据。总结起来：

> 正是由于使用了像线缆解调器这种允许用户相互窃听的通信技术，导致Internet不安全。

31.3 缺乏安全性是致命的缺陷

有些情况下，Internet的不安全性只是带来了一些不便，但在另一些情况下，这些问题会构成严重的威胁。例如，某个人获取了你发送的所有消息的副本，如果他公布了你与合作伙伴之间往来通信的内容，会导致这些商业机密的泄露；如果他在其中发现了你的信用卡或银行账号，会给你带来巨大的经济损失。

31.4 认证和保密是基本问题

上面谈到的安全问题都属于保密性问题（privacy problem），总结起来：

> 保密性是指对数据进行加密，如果只有发送和目标接收方能够获得消息的内容，那么这样的通信是保密的。

当我们需要在Internet上进行交易时，就会产生另一个安全问题：如何身份认证问题（authentication problem），这是身体鉴定的核心问题。我们通过一个简单的例子说明身份认证的困难性。假如某个ISP允许用户注册一个E-mail地址，出于玩笑，某人选择了一个明星的名字，那么他很可能会因此收到很多意想不到的邮件，这些信可能来自世界上最富有的人、最高法院的大法官，甚至是美国总统。

尽管上面这个例子显得很滑稽，但却揭示出身份认证并不容易实现的事实。又例如，某个用户在网上看到某公司的产品广告，为了购买产品，用户需要提供信用卡账号，但是用户怎样证实网站和广告的真实性？他访问的网页也许是某些不怀好意的人制作的假网站，它们和真网站一模一样，并等待受害者上当受骗（发送他们的信用卡账号）[6]。将问题总结起来：

> 认证是指确认通信对方身份的问题，客户在下订单之前必须要证实广告和公司信息的真实性。

31.5　数据可以被篡改

除了认证和保密之外，还有很多其他安全问题。Internet 上传输的数据需要经过很多不同的网络和路由器，那么接收方怎样验证收到的数据没有被篡改？例如，一个客户发送了一张网上购物订单，如果订单被第三方截获，并篡改了发货地址，这样客户必定收不到商品，而作为商家又怎样确保订单上的信息与客户本来填写的内容一致？

尽管绝大多数 ISP 的信誉很好，不会篡改数据，但我们始终不能排除这种可能性。因此，有必要使用一些技术来保证 Internet 上的数据不被人为改变。

31.6　编码可以保证消息的私密性

很久以前，人们就开始使用重新编码的方式对消息进行加密，例如，国王使用重新编码的消息与军队进行通信，只有发送方和接收方明白消息的真正含义，这样即使消息在中途被意外截获，内容也不会泄露。

现代计算机系统采用了相同的原理来保证信息的私密性。在通过网络发送消息之前，发送方计算机上的软件对消息进行编码，当编码后的消息到达接收方计算机后，接收方重复同样的动作获得原始消息。假设这个过程中使用的编码方式足够复杂，即使消息被第三方截获，他也不能够破译出原文内容。

31.7　计算机的密码使用了数学

目前，密码的破译手段已经今非昔比，由计算机取代了手工方式破译密码。一台计算机每秒可以尝试成千上万个组合，如果让多台计算机同时工作，所需的时间还将大大缩短。使用计算机可以轻而易举的破解采用过去手工方式加密的内容，因此 Internet 不可能沿用过去的加密方式，取而代之的是更复杂，更难于破解的数学编码方式。我们称重新编码的过程为加密，称解码的过程为解密。

那么加密解密和数学有什么关系呢？答案很简单：在数字计算机内部，所有的信息都是用数字来表示的，即使是"abcdef"这样的字符串也不例外。所以对信息加密的过程就是对数字的特殊运算，因此要使用数学。例如，我们可以使用字母替换的方式对消息进行重新编码（例如，用字母 x 替换所有的字母 e），对计算机来说，就是用表示字母 x 的数字替换表示字母 e 的数字。当然计算机网络中所使用的加密方法要比这种简单的替换复杂得多，所以需要更为高深的数学变换。

31.8　没有绝对安全的网络

采用基于数学的加密算法就可以高枕无忧了吗？当然不是，正如一把锁换不来 100% 的保险

---

[6] 实际上，很多骗局只是利用了一些用户粗心大意的毛病，假冒网站和真实网站的名称往往只有一字之差，如果用户不小心键错了网站名称，便可能会访问到这些假冒站点。

一样，加密也不能带来绝对的保密：只要第三方有足够多的计算机，花足够长的时间，就能成功地进行破解密码。然而，通过谨慎选择加密算法，我们还是可以保证加密的内容在相当长的时间内不会被破译，从而取得相对的安全。总结起来：

> 尽管任何网络都不是绝对安全，但现代密码学已经能够做到即使采用最先进的计算机来破解密码也要几年时间。

在谈论 Internet 安全时必须要牢记上述原则，如果有人声称某种加密算法非常安全，这并不代表它不可破解，只是破解它很困难。因此当我们说一种算法比另一种算法更安全时，是指破解它需要更长的时间。

31.9　加密保障了电子邮件的私密性

加密可以保障电子邮件的私密性，用户输入消息内容后，由特定软件对其进行加密，然后通过 Internet 发送。因为内容已经加密，且除接收方之外没有其他人能够解密，所以发送方不用担心信息泄露。图 31-1 阐释了这其中的原因。

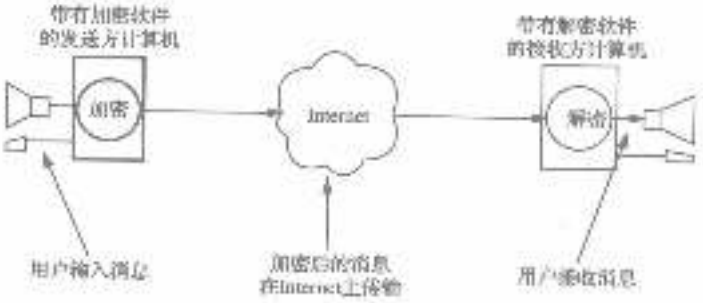


图 31-1　加密软件可以保证邮件内容的私密性，Internet 上传输的都是经过加密的消息

31.10　加密软件需要密钥

前面我们只是提到了解密工作是由特定软件来完成，假如某人使用同样的解密软件并设法获得了加密后的信息，是不是就可以成功破解原文了呢？答案当然还是否定的，为了说明问题，我们举一个生活中很常见的例子：汽车厂商生产了很多同一型号的汽车，每辆车都有一把独一无二的钥匙，其一辆车的钥匙不能打开其他车了。

加密软件与此类似，每个用户都有一把独一无二的钥匙：密钥（key），只有拥有了这把钥匙，才能解密发送给该用户的信息。

31.11　两把密钥意味着不需要预先信任关系

在早期的计算机加密方案中，每个用户只有一个密钥，这意味着，消息的发送方用这个密钥进行加密，而消息的接收方必须使用同样的密钥才能解密。今天，Internet 上使用的绝大多数加密系统都为每个用户提供两个密钥，一个用于加密，另一个用于解密。用于解密的密钥不能公开，因此称之为*私钥*（private key），用于加密的密钥可以公开，因此称之为*公钥*（public key）。这种使用两个密钥的加密技术总称为*公钥加密技术*（public key encryption）。

为什么公钥加密技术能够获得如此广泛的应用呢？主要原因是这种技术可以让任意两个人，在没有预先信任关系（共享密钥）的前提下，进行有加密保护的通信。我们可以通过一个客户与公司在线交互的例子来解释"预先信任关系"的含义。如果采用传统的单一密钥加密系统，要想

让所有客户都能进行加密通信，每一个客户都必须拥有该系统的密钥。而且客户只要获得该密钥，就可以解密其他人所发送的加密消息，这样一来会给某些居心不良的人带来可乘之机。因为公司不可能与所有客户都建立有“预先信任关系”，就不能排除恶意用户的存在，因而就不能保证这种加密系统的可靠性。如果使用基于公私密钥对的加密系统，情况就不同了：公司可以让所有客户都获得其公钥，但保证其私钥的机密性，这样每个客户在向公司发送信息时用公钥加密，只有公司能够解密。因此我们说双方在没有预先信任关系的前提下实现了加密通信，换句话说：

使用公钥加密技术后，通信双方可以不再依赖预先信任关系。

为了说明公钥加密技术的使用方法，假设你想要接收加密邮件，首先要获得一个公私密钥对和需要的软件。通常加密密钥的长度在 56 位到 256 位之间（相当于 17 位到 80 位十进制数字）。简单起见，我们假设私钥为：98989898，公钥为：35353535。其中私钥需要秘密保存，不能让别人知道，公钥则需要对外公布，你可以直接告诉朋友或发布到专门的列表中[○]：

John Smith 的公钥是 35353535

密钥的数学性质保证了不可能从公钥推算出私钥。当有人想给你发送加密消息时，他们首先要通过某种途径获得你的公钥，然后用你的公钥来加密消息，因为只有你才有对应的私钥，所以别人无法解密。

你的公钥只能用来加密发送给你的消息，而私钥只能用来解密这些消息。如果想给别人发送加密消息，你必须拥有他的公钥，而对方再用他的私钥去解密。这样一来，你和他之间要实现双向加密通信，至少需要 4 个密钥：你的两个密钥用于处理发送给你的所有消息，而他的两个密钥则用来处理发送给他的所有消息。图 31-2 说明了这个过程。

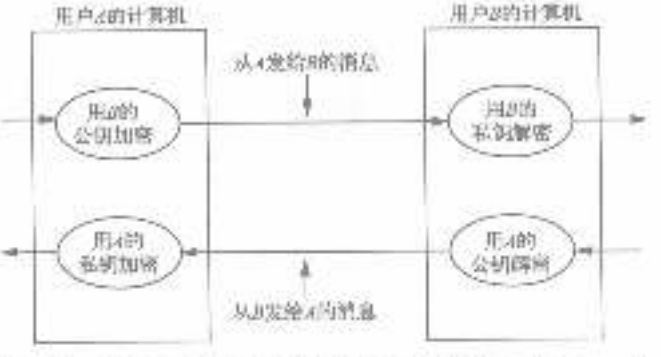


图 31-2　本例有两个用户使用密钥进行加密通信。在 Internet 上传输的所有消息都经过加密，但是根据传输的方向不同，使用的密钥也会不同

31.12　安全电子邮件的实际应用

怎样将上面的想法应用到实际中去呢？如果用户想加密消息，他应该选择什么软件？虽然市场上有很多加密软件，但还没有哪一种加密技术拥有统治地位。在电子邮件应用领域，就存在了两种流行的加密系统。安全的多用途网际邮件扩充协议（Secure Multipurpose Internet Mail Extension，SMIME）是在现有的邮件标准 MIME 的基础上增加了安全功能。它的优势在于良好的兼容性：不要求改变邮件的基本格式。另一种可以进行加密邮件通信的系统是 PGP（Pretty Good Privacy），最早是由一个 MIT 的学生，为了填补这个空白而开发的。虽然市场上很多邮件软件同时支持这两种加密系统，但 PGP 还是暴露出不能与 MIME 之类的邮件标准兼容的弊端。

31.13　安全的无线网络

根据第 19 章的介绍，目前市场上已经有一些便宜的无线路由器可以通过 NAT 技术允许家

○ 已经有了专门提供公钥发布服务的公司。

庭或公司中的多台计算机同时上网。由于不需要物理连接，无线网络的安全隐患更多，问题也更突出。一方面，隐患体现在对通信保密性的威胁：在无线路由器覆盖范围内的计算机可以在不被察觉的情况下窃听其他无线用户的通信内容。另一种安全威胁是被动共享（Inadvertent sharing）：无线路由器覆盖范围内的所有计算机都可以通过其免费访问 Internet，这种行为会违反用户与 ISP 签订的服务协议，即便 ISP 没有要求，那些不速之客也会影响可用带宽的大小。

还可以安全使用无线网络吗？可以，目前已经有很多方法可以实现保密通信和访问控制。例如，为了保证私密性，无线局域网（Wi-Fi network）提供了一种有线等效协议（Wired Equivalent Privacy），简称 WEP。使用 WEP 后，计算机与无线网关之间的通信内容都获得了加密保护，并阻止第三方窃取。此外，无线局域网还引入了一种称为 SSID 的标识符（SSID 的全称为 Service Set IDentifier），只有知道 SSID 的计算机才能通过无线网关上网，并收发数据包。

这些无线网络中的安全措施需要正确配置，也就是说网络的所有者在建立无线网络时需要决定是否启用 WEP 加密、是否使用 SSID、所使用的 SSID 是广播形式通知给所有计算机还是秘密持有。要点是：

> 尽管无线网络提供了加密和 SSID 等机制来解决通信私密性和未经授权的访问，但它们只有经过正确配置后才能生效。

31.14 防火墙可以阻止不想要的数据包

除了上面提到的安全问题，计算机或网络也会收到一些不受欢迎的数据包。例如，攻击者会发送一些数据包探测目标计算机是否开启了某些服务，如 Web 服务、ftp 服务或远程桌面。一旦攻击者发现某个服务在运行，便可针对该服务发起攻击（例如，猜测用户和密码）。

最常用来保护内网不受外界攻击的技术是 Internet 防火墙（Internet firewall）。防火墙一词原本是一种防止火势蔓延，起到隔离作用的消防设施。Internet 防火墙放置在被保护的网络与外网之间。图 31-5 说明了防火墙的位置。

图 31-5 防火墙可以保护一个或一组计算机免受外界攻击

31.15 防火墙可以对数据包进行过滤

防火墙是怎样工作的呢？在投入使用之前，网络管理员根据需要的安全策略对防火墙进行配置。使用过程中，防火墙检测每一个数据包，只有符合安全规定的数据包才能通过。例如，某个网站没有对外提供 FTP 服务，那么防火墙就会拒绝所有 FTP 访问请求。也可以通过配置防火墙允许访问网站上某些主机上的 FTP 服务，而拒绝访问其他的主机的 FTP 的请求。

防火墙的作用体现在哪里？主要是防止意外访问。例如，网络内部的某个员工因为误操作而启动了计算机上的 FTP 服务，这时防火墙就可以阻止不希望的 FTP 连接。更重要的是，防火墙只允许某些已知的访问通过，会阻止对未知服务端口的访问，这样便可以抵御一些未知的攻击。

31.16 防火墙可以防止特洛伊木马

有时候员工会发现企业的防火墙不仅可以限制进入内网的数据包，同样也可以控制从内网向 Internet 的访问。为什么需要如此设计？因为如果企业不能控制向外的连接，便很容易成为特洛伊木马的攻击对象。只要内部网络中某个员工不小心运行了木马程序（通常掩饰成游戏或动

画，或是作为邮件附件）就会中招。通常，特洛伊木马通过邮件传送给员工。表面上看可能是播放了一段动画，实际上却在连接外网的某个计算机，并允许他们访问和控制这台内网中的计算机。如果正确配置了防火墙，就可以阻止木马程序向外连接。总结起来：

> 防火墙是放置在内网和Internet之间的一种网络设备，可以控制外界对内网的访问。为了防止特洛伊木马，防火墙也控制着公司内网向外发起的连接。

31.17　住宅和个人防火墙

前面介绍过可以在企业网络中部署防火墙来保证公司计算机和网络安全，其实防火墙还可以保护住宅小区中的网络和计算机，这种防火墙一般有两种形式：硬件设备和软件模块。

硬件设备（hardware device）。实现全部功能的防火墙设备可以做的很小，价格也很低，这样，单个用户可以购买一个独立的设备放在计算机与ISP之间（如计算机和DSL或Cable Modem之间）。防火墙会阻止不满足条件的数据包。例如，第18章中介绍的无线路由器（即NAT盒）通常都内置了防火墙功能。

软件模块（software module）。除了选择硬件设备，还可以在计算机上运行一个具有防火墙功能的软件。这类软件位于操作系统和网络接口之间，保证所有进程计算机的数据包都要经过防火墙软件。防火墙软件与其他软件明显不同，在功能上与独立的硬件设备功能相当。

应该使用硬件防火墙还是软件防火墙呢？硬件防火墙可以处理一个小区之内所有计算机的数据包，因此可以包含共享这个ISP连接的所有计算机。而且，还可以在其他设备（带有NAT功能的无线路由器）中嵌入防火墙功能。相反，软件防火墙就只能保护它所在的计算机，但它也有自己的优势。软件的安装更简单，而且，方便用户进行设置（例如，修改防火墙规则）。因此，很多用户同时使用这两种形式：无线路由器上的防火墙和自己计算机上的防火墙软件。总结起来：

> 防火墙不仅可以用来保护企业网络，还可以保护小区中的网络和计算机。用户既可以选择能够保护整个网络的硬件设备，也可以选择只能保护本机的防火墙软件。很多用户会同时使用这两种防火墙。

31.18　有一类程序可以用来检测黑客入侵

除防火墙之外，大的公司网络中往往还部署了入侵检测系统（Intrusion Detection System），简称IDS。IDS可以动态监测网络中的异常行为。即使网站已经在防火墙的保护之下，IDS仍然有必要。原因很简单，随着笔记本电脑和移动存储设备（如FlashROM）的广泛使用，使得恶意软件偶然出现的概率大大增加，IDS的作用就体现在一旦它检测到这种情况，就会通知相关人员采取相应措施。

31.19　服务可能被拒绝

还有另外一种攻击会阻碍正常通信，网络专家用分布式拒绝服务攻击（Distributed Denial Of Service attack）来表示这种类型的攻击，简称DDOS，并读作“d dos”。DDOS攻击以破坏网站或组织正常连接和使用Internet的能力为目的，这种攻击很类似用封死商场的入口的方式，来阻止顾客进入商场购物——攻击者调动Internet上很多主机同时向目标站点发送数据。如果主机足够多，发送速度足够快，很快大量数据就会阻塞目的站点，就如同公路上发生的堵车情况一样。

DDOS很难检测，多数DDOS攻击的目标非是规模较大的组织而非个人。还好，这些组织都有专门人员与ISP协作来定位攻击源头，控制攻击，但是目前针对这一问题，还没有理想的解决

方案，总结起来：

> 尽管入侵检测系统可以发现网络中出现的大量未知数据包，但却很难阻止由此造成的分布式拒绝服务攻击。

31.20 小结

尽管所有的网络都不是绝对安全的，但现代加密技术已经可以充分保证数据在Internet上传播时高度的安全性，阻止第三方篡改或解码数据。公钥加密技术则是这种安全性的基石，用户获得一对公私密钥（私钥保密，公钥公开），发送方用对方的公钥加密消息，接收方用私钥进行解密。

带有NAT功能的无线路由器和无线网络面临着很多安全隐患，通过对无线网络设备的合理配置可以保证无线通信的私密性并防止无授权用户的非法访问。

防火墙可以检测进出网络的数据包，一方面，它可以阻止不需要的数据包进入网络或计算机，另一方面，它可以防止特洛伊木马程序在用户不知情的情况下向外传输数据。入侵检测系统则可以进一步发现网络中的异常情况，但两者都不能阻止分布式拒绝服务攻击。

第 32 章　远程安全接入（VPN）

32.1　引言

第 31 章介绍了一些保证网络安全的技术。本章将继续讨论这个话题，并介绍另一种网络安全技术，利用这种技术用户可以像访问本地网络一样方便、安全地远程访问企业网络，并解释这种技术的工作原理。

32.2　组织机构中的员工享有特殊的权限

要理解与安全远程访问相关的问题，必须先了解一下企业网络如何保证安全。问题的关键在于大多数企业网络对内网主机和外网主机有着清晰而严格的界限。内网主机由企业自己的员工所使用，外网主机则包括供应商、客户，甚至 Internet 上任何人或企业组织所使用的计算机。

企业首先制定访问控制策略，这些策略分别规定了内网主机和外网主机可以做什么和不可以做什么，然后采用一些特定的技术来上述策略付诸实施。一般来说，内网主机拥有更多权限，例如访问企业员工数据（如邮件地址、电话号码、办公地点等），相反，外网主机则没有这方面权限。

第 31 章介绍的防火墙（firewall）就是一种实现这类内外有别的策略的技术。部署在内外网结合部位的防火墙能够判断那些数据包符合策略，并允许其进出。要点是：

> 内网用户通常比外网用户拥有更多的权限。为了体现这种差别，企业网络中采用防火墙来阻止来自外网对内网的某些访问。

32.3　外出的员工不再有特殊的权限

虽然强调内网用户和外网用户的差别是十分重要，但在实践过程会造成很多问题。例如，某个员工到外地出差，他在访问和使用企业网络时，自然成了外网用户，这就使得员工降级成为外网用户。

注意，这里提到的内网和外网的定义是针对物理位置而言，不能因为出差员工使用的是同一台笔记本电脑就认为他仍然是内网用户，只要计算机通过外地网络连接 Internet，就是外网用户。总结起来：

> 当员工不在企业内部时，他就成为了外网用户，因而失掉了其特权，这一点与其所使用的计算机无关。

32.4　远程用户没有特殊的权限

当员工暂时在企业外部工作时，其对内网的访问权限也随之消失，这个问题对于始终在企业外部工作的员工来说就更突出，我们称此类用户为远程用户（telecommuter）。

远程用户从当地的 ISP 接入 Internet，这种方式对于浏览网页，登录企业网站等常规任务没有影响，但是当他们访问企业内网时，必然因为其外网用户的身份而导致失败。要点是：

> 如果简单地剥夺所有外网用户对内网的访问权限，那么企业的远程用户就无

定无法使用内网资源。

32.5 租用专线可以保证通信安全

电信公司提供了一种解决远程用户访问问题的方案。利用他们提供的线路租用服务，企业可以指定任意两个地点，并租用一条连接双方的线路。而且，只有该企业能够使用这条租用的线路（只有位于线路两端的设备，才能访问线路上的数据）。因此，通过线路租用方式可以获得并建立专用网络连接。

为了支持远程用户访问，企业首先要租用一条连接远程用户所在地（例如，自己家）与公司的专线，并在线路两端分别安装需要的网络设备。在远程用户一端将计算机与线路连接；在企业一端让线路与内部网络相连。这种方法在本质上是通过租用的线路将企业内网的范围扩大至远程用户的所在地，因此远程用户就获得了使用内网的权限。图32-1说明了这种连接方式。

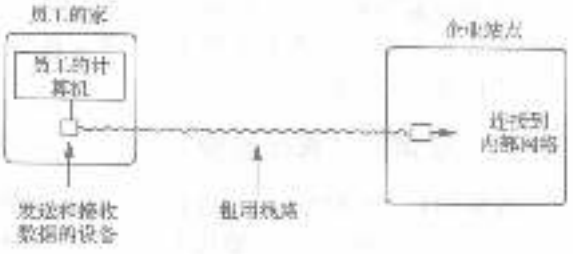


图32-1 租用线路将远程用户的所在地和企业连接起来，这种方式允许远程用户安全的使用内网资源

尽管租用线路是一种可行的方案，但它有一个缺点：价格昂贵。即使租用一条低速线路每月也要几百美元，况且，这种方法还不能解决员工出差期间使用内网的问题。所以需要有更好的解决方案。

32.6 标准的Internet连接更便宜

第14章介绍过，DSL和有线电视网络允许网络运营商提供价格低廉的宽带接入服务。其带宽可以满足远程用户的需要（收发电子邮件、文件传输和远程桌面等）。

宽带接入的方式对出差人员同样有效，很多宾馆都提供这种服务，有些地方甚至是免费的。

32.7 能否兼顾二者之长

不过宽带接入还不能彻底解决问题，尽管它价格便宜且性能优良，但用户通过它访问企业内网时仍然被视为外网用户。为了让工作在企业外部的员工都能使用内网资源，他们必须获得和内网用户一样的权限。问题是：能否同时获得租用线路的优点和宽带接入价格优势？也就是说，能否找到一种技术让远程用户低成本且无限制地使用内部网络。

32.8 虚拟专用网是一个解决方案

聪明的工程师们设计了一种同时具备低成本和高安全特性的技术，并利用它很好的解决了上述难题。使用这种技术，远程计算机也可以获得全部权限，使得用户如同在企业内部使用网络一样。而且，这种技术很安全，即使在公共场所（如宾馆）使用无线网络也能保证数据安全。换句话说，如果第三方截获了通信数据，他既不能破译通信内容也无法获得访问权限。

网络业界给这种技术起了一个形象的名称：虚拟专用网（Virtual Private Network）。说它专用（private）是因为只有发送方和接收方能够解密数据包中的内容。说它虚拟（virtual）是因为用户并不需要真正租用线路。因为这个名字很长，通常简称为VPN。总结起来：

通过一条加密的Internet连接，VPN技术可以让远程用户安全、廉价且不受

非常的使用内网。

32.9　VPN 怎样工作

不难想像，加密技术是 VPN 的基石。为了保证私密性，VPN 加密每一个数据包。也就是说，数据包在发送之前要先经过加密，然后再传输，最终在目的地进行解密。这样即便所有数据包都是在 Internet 上传输，也能够保证其私密性。图 32-2 展示了 VPN 各个主要模块的工作原理。

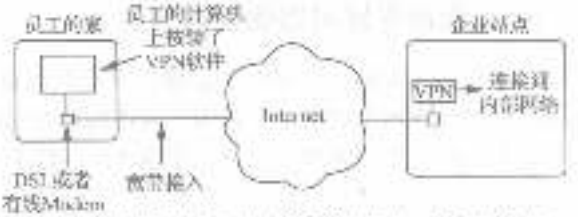


图 32-2　实现员工远程访问所需的模块。员工计算机和企业内部网络都需要运行 VPN 软件

32.10　所谓的“直接连接”

尽管加密技术解决了远程连接过程中的保密性问题，VPN 还需另一种技术实现所谓的“虚拟”网络连接。外出员工计算机上的 VPN 软件必须要制造一个假象，让机器上的其他网络应用感觉好像是直接连接在企业内部网络上。它们发出的每一个数据包（数据报）的 IP 地址都必须是内部网络的一个合法地址。为了制造这样一个所谓的“直接连接”，VPN 软件必须要切断其他网络应用与物理连接的关系，并取而代之获得对物理连接的惟一控制权。启动之后，VPN 软件与位于企业内部网络的 VPN 软件进行通信，并获取一个合法的内网 IP 地址给操作系统。只有 VPN 软件有权使用真正的 Internet 连接（即物理连接）和对应的外网 IP 地址，而主机上的其他网络应用只能使用 VPN 软件获得的内网 IP 地址，所以从应用程序的角度来看，计算机就好像直接连接在企业内部网络上，因为它们收发的数据包使用的都是内网 IP 地址。

如果远程计算机上的网络应用产生的数据包使用了企业内网 IP 地址，那么这些数据包怎样正确路由呢？实际上，它们并没有被直接发送出去。VPN 软件将每一个这样的“内网”数据包放入另外一个数据包之中，包装后的数据包可以正确路由到企业网络，届时，企业网络上的 VPN 软件会剥开这层外壳，向网络内部转发包装前的数据包，这样，这些数据包就好像是从本地发出的一样。

为了更好地理解 VPN 的工作流程，假设你在芝加哥，并想写一个朋友通信，但希望让他误以为你住在费城。为了让他相信，你的信上必须有费城的邮戳，而且回信地址也在费城。如果你有一个亲戚在费城，那么这个小骗局就很容易实现。首先，你在信封上注明朋友的收信地址，并且用那个亲戚的地址作为回信地址，然后将这封信放到一个更大的信封之中，把这个大信封寄给费城的亲戚。当亲戚收到信后，取出其中的小信封并寄出。同样，当你的亲戚收到朋友给你的回信时，将回信也放入一个更大的信封中寄回给你。

一旦明确了这个流程，你就可以让更多的朋友上当受骗。有趣的是，你的亲戚并不需要事先知道你要联系的人，也不需要知道写信给你的人。他要做的就是，收到你寄出的信时，打开信封，把里面的信直接投递出去，或者收到给你的信时，把它们放在一个更大的信封中，寄给你在芝加哥的地址就可以了。

VPN 采用的是同样的方法：它将一个数据包放入另一个数据包之中，来欺骗接收方，让他们认为远程用户也在本地网络上。图 32-3 描述了一个数据包在经过 VPN 各个处理阶段时的不同变化。

a) b) c) d) e)

图 32-3 数据包在经过 VPN 处理时的变化情况。a）原始数据包；b）经过发送方 VPN 软件加密；c）加密后的数据包被放入另一个数据包中用于实际传输；d）在接收方，外层数据包被 VPN 软件除去；e）解密获得原始数据包的内容，并在内部网络中进行转发

32.11 VPN 的重要性

对于很多 Internet 用户而言，VPN 技术改变了他们使用 Internet 的方法。出差在外的员工可以如同在企业内部一样不受限制的使用内部网络。更重要的是，他们不需要再担心安全问题。VPN 保证了通信的安全性。因此，企业员工在任何地方（客户，甚至竞争对手）都可以通过网络及时获得企业内部的各种资源。要点是：

员工即使不在企业内部也能够通过 VPN 技术不受限制的访问内网资源。由于 VPN 良好的安全性，它改变了非差员工使用 Internet 的方式。

第 33 章　Internet 经济和电子商务

33.1　引言

本章将从经济角度来讨论 Internet。首先介绍相关基本术语，解释为什么可以在 Internet 上开展贸易。进一步扩展前面几章有关安全性的讨论，分析加密技术如何确保在线交易的安全。

33.2　谁为 Internet 买单

要讨论与 Internet 相关的经济问题，就不能不谈如何盈利的问题。谁来投资基础设施？当然，个人用户要付费才能上网，费用的多少取决于用户和 ISP 之间链路带宽的大小。每个客户与 ISP 签订一份服务等级约定（Service Level Agreement），简称 SLA。SLA 中除了规定了客户需要支付的费用外，还明确规定了服务质量水平，例如，链路的带宽和每月总流量等。面向企业用户的 SLA 又比普通居民用户的 SLA 更复杂，体现在对出现问题后响应时间的保证和超出流量限制部分的收费方式上。

本地（第 3 级）ISP 是否需要支付网络使用费？答案是肯定的，3 级 ISP 需要依赖 2 级 ISP 提供的网络并根据流量支付使用费。类似地，2 级 ISP 又是 1 级 ISP 的客户，必须为其提供的服务支付费用。因此，低级别的 ISP 需要从用户那里获得足够的收入，才能维持收支平衡。

那么 1 级 ISP 的情况又是如何呢？他们也是其他 ISP 的客户吗？一般来说，并不是这样，1 级 ISP 之间地位平等，因此他们之间的合同也称为对等协议（peering agreement）。如果两者相互使用的流量相同，相应的费用也可以抵消。如果某月其中一方发送的流量超过另一方，对于多余的流量，双方会根据对等协议进行结算。总之：

> 居民和企业用户使用 Internet 需要向 ISP 支付费用。低级 ISP 需要向更高级的 ISP 付费。在 Internet 的主干网上，1 级 ISP 之间地位平等，互相提供服务，并只对多余的部分收费。

对于企业来说，使用并提供 Internet 服务是日常运营成本的一部分，这也就意味着企业会提高最终产品或服务的价格来弥补成本。

33.3　电子商务是一块大蛋糕

电子商务（electronic commerce，简写为 e-commerce）是指通过 Internet 进行业务交易，交易可以是面向客户的（如在线银行、在线投资和在线支付等），也可以是企业之间的贸易。从概念上看，电子商务与现有的交易方式（如电话银行、邮购、传真订单等）没有本质区别，差异主要体现在细节上。

对客户而言，电子商务的直观感受就是在线订货。客户在网站上浏览、选择要购买的产品，再选择一种支付方式就可以完成一个订单，这个过程中的每个环节都通过网络完成，无须其他物理介质（纸介质、电话等）。最终，不需要像支票这些纸介质进行付款，电子商务允许用户用电子方式付款。

从 1993 年到 2003 年的十年间，电子商务从一种新兴事务变为主流贸易形式。1993 年时，

拥有网站的企业还很少，能够通过 Internet 下订单的企业更是凤毛麟角。十年之后，很多企业无论大小都有自己的网站，而且在线下订单也非常普遍。与此同时，很多银行开通了网上银行，在线支付形式迅速流行。更显著的变化体现在 Internet 上完成的交易总额在 1997 年之后有了大幅度提高。根据 www.thestandard.com/metrics 上的一篇文章，仅在 2005 年 11 月 25 日到 12 月 27 日之间，12 万家在线零售商通过 VeriSign 完成的交易总额突破了 88 亿美元。

33.4 安全技术使得电子商务成为可能

面向个人用户的零售业务是电子商务的主体。普通的零售交易网站以目录形式列出所销售的商品，用户首先通过 Web 浏览器搜索感兴趣的产品，然后到交易网站上获得商品的详细信息。通常，网页提供一个独立的框架来显示已购物品信息，这个功能也被形象地称为购物车（shopping cart），用户可以操控购物车：创建新的购物车，向车内添加新商品，察看车内商品和价格，取消车内商品等。当用户完成需要商品的选择后，就可以真正购买了，填写发货信息以及支付方式后（如信用卡号），交易就自动完成了。

用户知道他们的信用卡号必须保密，如果 Web 浏览器不能保证这样的信息安全，也就不会有人愿意在线购物。因此，提供保密机制的浏览器就成了推行电子商务的前提。

33.5 安全套接字

浏览器怎样保证消息的私密性？答案还是第 31 章所介绍的加密技术。当网页上出现表单时，用户必须向其中填入信息，这些信息会通过 Internet 被发送到服务器一端，这个过程由浏览器完成。浏览器有两种选择：自动提供安全保障或提出警告。自动安全保障需要加密，如果表单要求加密传输，浏览器会自动加密要发送的消息。如果表单没有指定加密传输，浏览器会自动弹出一个对话框，警告用户将要发送的表单信息没有安全保护，由用户决定是否同意继续发送[⊖]。如果用户同意，信息将以明文方式传输；如果用户不同意，浏览器将不会发送该信息。

浏览器加密消息的过程，对用户来说是透明的。目前最常用的自动安全保障措施是安全套接字层（Secure Socket Layer，SSL）。使用 SSL，浏览器可以通过加密保证消息的私密性。更重要的是，这个过程都是由浏览器自动完成的，不需要用户参与。例如在访问前缀为 https://（而不是标准的 http）的 URL 时，浏览器会自动启用 SSL。为了确定连接是否加密，用户可以查看浏览器上的一个状态指示图标（一把处于打开或关闭状态的小锁）。

33.6 公钥加密技术提供了认证手段

安全技术与浏览器相结合更多用于加密信息。在进行加密之前，浏览器需要获得加密使用的密钥（如服务器使用的公钥），并验证服务器和密钥的真实性。所有步骤都是自动完成，没有对话框显示，并且用户不需要输入任何信息，这个过程完全由浏览器通过 Internet 自动完成。

出乎意料的是，浏览器也是利用公钥加密技术来进行真实性验证。要点是：

公钥加密技术应用广泛——不仅可以进行加密，还可以进行认证。

那么认证是怎样实现的呢？理解这一点需要了解公钥加密算法的数学特性。由公钥加密的信息只能由私钥解密，反之亦然，由私钥加密的信息只能由公钥解密。

当然，使用私钥加密的信息不能获得保密性，因为所有人都可以知道公钥。不过，如果你收到了经过某用户的私钥加密的信息，由于只有该用户自己知道他的私钥，因此可以保证该消息

⊖ 对安全性不很介意的用户可以通过浏览器设置避免弹出对话框。

的确是由该用户发出的。更重要的一点是，利用该用户的公钥，我们可以很容易地验证某个消息是否经过该用户的私钥加密。

如果你对此还不是很清楚，这也不要紧，毕竟公钥加密技术涉及很多数学的内容，很难一下子明白。而且，上面的讨论还没有涉及二次加密等更为复杂的密码技术。你只需记住，当你在网上购物或使用网上银行时，浏览器会利用公钥加密技术对所访问的网站进行认证，并加密整个会话。

33.7　数字签名

一种提供认证的机制称为数字签名（digital signature）。之所以用这个名称是因为这种用来识别电子文档真正发送方的方法与通过签名来识别普通文档的发出人的方法很类似。数字签名本质上是一段加密信息，由发送方的私钥加密而成。很多人误以为数字签名就是将普通签名扫描进计算机后所获得的图像，这只能称之为数字化的签名（digitized signature）。二者最大的差别在于数字化的签名很容易仿造假冒，而数字签名却不行。虽然我们在此不深究数字签名的工作原理，但总结如下：

数字签名包含一段特别的加密信息，使用的加密技术保证数字签名不能被假冒。

33.8　证书中的公钥

前面我们介绍过任何人都可以发布公钥，但是在浏览器认证数字签名之前，它必须先确认发送方的公钥。浏览器也不需要用户提供发送方的公钥，因为有专门的公司提供公钥发布服务。当浏览器需要获得或验证某个公钥时，它可以选择一个可信的服务公司，在获得需要的公钥后再验证签名的真实性，当然浏览器和公钥公司之间的通信也必须经过加密确保安全。

尽管我们说浏览器的目的是为了获取公钥，但它真正从公钥公司获得的是一张数字证书（digital certificate）。每张证书都包含了一把公钥和用于验证颁发该证书的可信公司的数字签名。

浏览器在需要的时候会自动请求数字证书，而不必通知用户。事实上，用户只能通过浏览器中的一些提示感知这一过程的存在，例如在联系主机的过程中，浏览器中可能会出现“获取证书中”的字样。

也有偶然的情况，浏览器可能会无法从可信公钥公司获得需要的证书。这时，浏览器会询问用户是否要尝试另一个证书源。对于涉及财务信息（如提交信用卡信息）的情况来说，不要使用其他证书源。

33.9　什么是电子货币

尽管还有很多与加密技术有关的细节没有被提到，但这已经提供了电子商务所需的一切条件：加密消息、认证网站并安全地获得电子证书。但有些用户还希望能够出现电子货币。

信用卡的方便体现在大宗消费上，缺点则是手续相对烦琐，需要用户输入账号等信息，还必须小心过期问题。另外，用户自己还要负责核对每月的账单。为了解决这些不便，很多组织都提出电子货币（digital money）或电子现金（digital cash）的概念。其初衷是：提供一种与人们日常所使用的现金功能相当，使用方便（类比信用卡更方便）的电子形式的金融工具，主要用于日常小额支付。

目前已经存在几种具体方案。我们可以将电子现金理解为一种特殊的借记卡（debit card）。用户首先通过网络银行授权将部分现金从其银行账号转存到电子钱包（digital wallet，相当于借记卡）中。成功后，银行返回一个 ID 作为电子钱包的标识符，用户可以将该标识符保存在计算

机中。当用户在网上购物时，可以使用电子钱包中的电子现金进行支付。当余额不足时用户可以随时到电子银行进行充值。

33.10 电子现金还未广泛应用

原理虽然简单，但实际操作过程中还需要处理很多环节。银行不可能将真正的现金放在电子钱包中，本质上，电子钱包是一段加密信息，其中包含开户行、账户信息和金额等。当一项交易需用支付时，交易必须通过电子钱包拥有者的认证，发送支付请求给银行并由银行完成转账。这个过程也称为微支付（micropayment）。微支付过程中的消息必须经过加密来保证安全，而且还要保证交易中涉及的商家、用户等信息的真实性。

因为使用电子现金涉及大量的加密技术，这增加了软件设计的难度。更重要的是，商家、用户和银行必须使用同样的软件才能实现这样一个交易过程，这增加了推广和使用电子现金的成本。还有一点，电子现金主要面向小额支付，利润空间不大，因此很多建立电子现金服务的尝试都无果而终，至今还没得到广泛应用。

在新的服务尝试还在酝酿，总的来说是尽量减少微支付过程中与银行的往来。有些方法是将微支付积少成多，月底一次性从用户的信用卡账户上结算，另一些方法将微支付与上网费绑定。这些新的尝试收效如何还要看市场反映。

33.11 企业和电子商务

前面的部分都是从用户的角度来讨论电子商务，包括从零售商的目录中挑选并购买个人物品。那么电子商务对企业究竟有什么样的影响呢？我们可以从企业内部和外部两个方面分析。当企业进行采购时，与个人客户没什么两样，电子商务可以替代传统的依赖邮件、传真、电话的采购模式，通常称这种业务为企业对企业（business to bussiness communication），简称 B2B。和前面个人电子商务一样，B2B 也采用加密技术来保障交易安全。

企业还可以利用电子通信手段处理内部事务。例如，某企业需要每月提供一份企业运行状况的总结报告，管理层根据报告情况调配资源。利用电子通信手段可以缩短报表的制作周期，减少工作量。又如，某企业利用无线网络监控和报告库存状况，并将结果保存在专门的服务器中。拥有多个外地分支机构的企业可以利用 VPN 降低各个分支部门间的通信成本。VPN 系统利用 Internet 提供安全可靠且价格低廉的通信服务。这些都有利于快速收集各个部门的运行数据，帮助决策层及时而全面地掌握企业运行状况。

33.12 关于税收和网络内容中立性

只要有涉及钱的地方就有争议，Internet 也不例外。政府在考虑怎样对 Internet 相关的经营活动收税，独立的服务提供商在考虑怎样将利润最大化。政府的税收主要体现在两个方面：对提供 Internet 服务本身征收的税款和对通过 Internet 进行的商品和服务买卖所征收的税款，即通常所说的销售税（sales tax）或增值税（value added tax）。就 Internet 服务业务来说，大多数运营商都处在监管范围之内，因为通信服务必须纳税，而且这些成本最终会均摊到用户身上。

更激烈的争论出现在 Internet 服务的收费方式上。为了更好地理解这一问题，有必要了解一下其历史成因。在 Internet 建立初期，电信公司的主要业务是用模拟信号设备提供语音服务，那时通信线路出租还只是作为辅助业务，因此线路的收费标准是根据带宽的大小来决定的，带宽越大价格越高。随着政府允许电信行业相互竞争，数字技术成本的不断下降，电信公司的的语音电话业务收入已不断下降。

到 2005 年，电话运营商普遍面临着一个巨大的挑战：像 Skype 和 Vonage 这样的公司已经开始利用 Internet 提供语音电话服务。[9] 为了增加收入，电信公司和一些大的 ISP 提出了一种新的计费方案，基于用户发送与接收的内容来进行流量计费，也就是说，像 Google、Vonnage、Skype 这些公司要为同样的网络流量支付更多的租金。

这项提议以及相关的讨论让消费者感到忧虑，原因主要是以下三个方面。第一，Internet 服务费用的上升会使很多小型企业不堪重负，最终导致只有那些大的富裕的公司才能使用 Internet。第二，ISP 会借此机会改变不同用户的服务质量，这样客户得到的服务质量会根据所选的 ISP 不同而不同。第三，企业会直接将增加的 Internet 服务成本转嫁给最终的用户，而 ISP 和各级运营商却在没有提高服务质量的前提下收取更多的费用。终端客户更倾向于网络内容中立性原则（net neutrality）。根据这一原则，使用 Internet 的费用只与流量大小有关，而与流量的内容和目的地无关。这里的要点是：

> 网络内容中立性原则是指一种计费方式，使用服务时产生费用的多少只与所使用流量的大小有关，而与流量的类型和目的地无关。用户希望继续使用这种计费方式，而运营商和大的 ISP 希望用新的计费方式取代它，并借此提高他们的营业额。

到本书截稿时止，美国政府仍然坚持中立性原则，但是大的 ISP 也依然在为破旧立新而努力游说政府。

---

[9] 第 28 章介绍了语音电话服务（VoIP）。

第 34 章　全球数字化图书馆

34.1　引言

前面几章介绍了很多 Internet 上的服务以及它们各自的作用。更重要的是，每一章都阐述了一种服务背后的基本概念，例如超媒体、网页搜索等。

本章将总结前面讨论的内容，解释数字图书馆的概念，以及为什么说 Internet 已经成为全球范围的数字图书馆。除此之外，本章还预测了 Internet 上可能出现的服务，这些服务之间怎样相互协作，以及 Internet 的发展趋势。

34.2　各种各样的服务

书中介绍了很多基本概念，它们是各种 Internet 服务的基础。其中包括用于用户之间联络的电子邮件，万维网使用的超媒体浏览，搜索引擎使用的自动搜索技术，以及即时消息、音频、视频流的传输和文件传输等概念。每章上面提到的概念涵盖了很多 Internet 上提供的服务和功能，但还不是全部。例如，电话会议（telconferencing），它允许一组用户参加在线会议，而且就像普通开会时一样，与会者可以看到彼此的图像，听到各自的声音。实际上，现有服务的数量已经很多，一个人要花费几个月的时间才可能全部掌握。总的来说：

> Internet 上提供了各种各样的服务。用户可以传输、浏览或查询信息，购买商品或者与其他用户交流。Internet 上的服务包罗了各种形式的数字化信息，如文字、声音、图片以及音频和视频等内容。

34.3　新服务总是有规律地出现

尽管已经有了很多种服务，Internet 还在不断发展。程序员和研究人员还在不断地开发新的服务，有些令人激动的服务问世还不到十年时间。如果不了解下面的事实，就不可能正确认识 Internet 的巨大价值：

> Internet 还在发展，研究人员还在不断地探索存储、传输、引用和访问信息的新方法。

Internet 不断变化的服务和功能，会产生两个重要影响。第一，任何所谓的“服务大全”都会很快过时。第二，Internet 用户总是有机会学习使用新的功能。

34.4　灵活性可以适应变化

为什么 Internet 上可以有种类如此繁多的服务，而其他通信系统却不能？答案很简单：Internet 所采用的技术具有更好的灵活性。

本书中介绍的所有服务都是建立在第 12 章到第 19 章介绍的 TCP/IP 协议基础之上的。TCP/IP 允许用户像发送电子邮件那样发送图像和声音。实际上，支持很多用户同时使用的大型计算机系统通常只有一个 TCP/IP 协议栈，由所有用户共享使用。更重要的是，很多使用 TCP/IP 协议的计算机和 Internet 服务都是在它出现之后才被设计开发的，这已经充分证明了 Internet

能够适应外界条件的改变，总的来说：

> Internet 所采用的底层通信技术具有非常好的灵活性。正是凭借这一点，才使得各种后续出现的计算机和服务都可以工作在 Internet 之上。

34.5　数字化图书馆

数字化图书馆（digital library）这个术语已经被广泛用来描述大量数字信息的仓库。与传统图书馆一样，数字化图书馆也是各类知识的仓库。与路边的报刊亭一样，数字化图书馆也能提供不断变化的信息。与电话电视一样，数字化图书馆也能够让人们第一时间获得各种消息。

因为信息能够以多种形式存在，所以数字化图书馆要囊括文字、声音、图像、照片、视频以及人物访谈等各种内容。此外，数字化图书馆还可以提供一些实时信息。例如，用户面部表情的变化情况、某个地区的天气情况、高速公路上的车流情况、或某个机场的航班延误情况等。

34.6　卡片目录和搜索工具

尽管很多图书馆都已经开始使用计算机，但传统图书馆还在使用卡片来记录图书和期刊信息。所有卡片放在一起就构成了卡片目录（card catalog）。卡片目录通常都是两套并用，一套按照作者排序，另一套根据主题排序。

图书馆的卡片目录是其最主要的索引机制。人们只需要通过卡片目录就可以找到需要的图书或期刊，而不必像大海捞针一样在书架中乱转。而且，可以从标题或者作者两方面信息找出需要的资源。

数字化图书馆则提供了更加多种多样，功能也更强大的索引和查询机制。计算机的应用使得创建文档索引以及在文档中查询各种信息成为可能。查询机制可简单也可复杂（会找包含某个单词的文档或者包含整个词组的文档）。这里的要点是：

> 与传统图书馆所使用的卡片目录不同，数字化图书馆采用了多种索引机制，并在此基础上提供了多种查询方法。因此，用户可以通过多种途径找到某一段信息。

34.7　Internet 上的服务可以整合

数字化图书馆中的查询机制可以被整合到现有的服务之中，从而形成服务之间的交叉引用。实际上，前面几章介绍的很多服务都可以整合到一起，从而让一种服务能够使用另一种服务所提供的信息。浏览器就是一个诠释服务整合的范例。虽然浏览器最初是设计用来显示万维网上的超媒体文件，但我们现在还可以用它访问 FTP 站点发送电子邮件、收听音频广播、与朋友分享照片，甚至是看电影。

有趣的是，整合有利于增加一项服务自身的价值。例如，最初用户只能用命令行方式使用 FTP，这样用户必须记住很多命令才能使用。而当 FTP 功能被整合到浏览器之中后，用户就可以通过图形界面来使用 FTP。浏览器会自动显示一个文件夹中的内容，用户不再需要手工输入“dir”命令；用户在浏览器中点击一个文件的名字就可以下载该文件，不再需要使用“get”命令。

另一个服务整合的例子是：有些网站提供电子邮件方式的文件访问服务。用户向一个特殊地址发送一封电子邮件，邮件的内容要包括用户所需文件的名称。当该消息到达目的地址后，服务器上的程序会自动获取所需的文件，并放在回复的邮件中。这种方式对偏爱电子邮件的人十分有用。这里体现的基本思想是：

尽管每个Internet服务都是独立设计的，但却可以通过某种方式将它们结合起来。一种服务可以访问其他服务提供的信息。

34.8 Dewey先生，你在哪

几个世纪以来，图书管理人员都在竭尽全力地整理人类创造的各种知识。书籍是应该根据书名排序呢，还是应该根据作者或者主题排序？图书馆应该将小说和其他内容分开存放还是放在一起管理呢？

Melvil Dewey提出了一种管理图书的方法。他发明了一种统一编码方案，将图书根据主题分类管理。这就是杜威十进分类法（Dewey Decimal System）。这种方案成功地建立了一种标准化的图书分类管理方法。

毫无疑问，建立在Internet之上的数字化图书馆也需要这样一套分类方法。尽管人们已经开发了很多工具，但是大量信息还是以随意的方式存在。例如，万维网中的内容，所有个人和组织都可以建立自己的网站，他们可以自行决定网页中的内容和显示方式。结果就是：每个网站都有自己的信息组织方法和导航方式，根本没有统一的标准。一个用户好不容易掌握了如何在某个公司的网站上查找需要的内容，当他需要使用另一个公司的网站时，可能又要有一个重新学习的过程。与此类似，每所大学的网站的组织方式也都各不相同。总而言之：

Internet上的每台计算机都可以用自己的方式管理和发布信息，但缺乏一个统一的标准。

34.9 数字化图书馆中的信息

我们一直都在讨论查找与访问数字化图书馆中的信息的方法，却忽略了这些信息本身。有人会问："究竟可以获得什么信息？"还有人会问："最有趣的信息在哪？"这些问题就等于在问"那本畅销小说是我最喜欢的？"或者是"报摊上那份报纸的内容最有趣？"

这些问题的答案取决提问者本人。而且，当他们想从数字化图书馆中找出答案时，问题就不是那么简单了，因为要找的信息总是在变化。例如，有些Internet上的讨论只持续了几分钟，有些则持续数天，还有些文档会存在几年。在数字化图书馆中，用户必须借助自动搜索工具来找出感兴趣的内容。他们只需要向搜索引擎提供一些关键字、词组或其他描述信息就能找到并浏览需要的信息。

34.10 Internet是什么

在这本书的开篇，我们提出了"Internet是什么？"这个问题。现在我们可以用一段话作为回答：

Internet是一个建立在十分灵活的通信技术之上的，非常成功而且在快速成长的全球数字化图书馆。Internet数字化图书馆提供了各种各样的服务用于创建、浏览、访问、搜索、查看和传送信息。这些信息涵盖了很多领域，从科学实验的最新成果到娱乐活动的讨论。Internet数字化图书馆之中的信息可以记录在备忘录中，可以组织为超媒体文档，或保存在文本文档中。此外，在数字化图书馆中可获取的信息还包括一些可以实时产生并传输的数据，例如声音和视频。这类数据无须保存到磁盘上。而且，各种服务都被整合在一起，可以互相调用。用户浏览的内容可以无缝地从一台计算机上的信息切换为另一台计算机上的信息，用户还可以从一种服务切换到另一种服务。

34.11　个人见解

几年前，在我工作的那所大学网络委员的一次会议上，图书馆的工作人员兴冲冲的描述了他们是如何安装服务器，以便让学生、教职员工以及校外人员能够通过 Internet 来访问图书馆中的资源。开始的时候，他们主要是介绍访问方法：电子邮件、新闻组和网页。

在倾听的过程中，我突然意识到一次巨大的变革正在发生。他们谈论的不仅是如何自动化卡片目录或者如何为本地文档创建索引。他们正在描述一个全新的未来：所有的信息都以数字形式来保存，会自动为它们建立索引，并且在需要时通过网络传播。他们还预言搜索引擎将会得到广泛应用。总之，图书馆的工作人员并不只是在尝试一些新的访问方法，他们更是在为一个数字化的未来做着准备工作。那时，现在意义上的图书馆将会消失，取而代之的是无处不在的信息网络，一切知识和信息都可以从网上获得。

Internet 术语表

尽管学习任何新的术语都很困难，但学习 Internet 术语尤其令人没有信心。对于初学者来说，用于描述 Internet 服务、网络技术以及特定计算机程序的术语几乎或者根本就弄不懂。Internet 术语表来自于计算机联网、商业、政府部门以及商业产品等各方面。

本词汇表对 Internet 中使用的术语作了简要的定义，着重介绍被广泛使用的术语，忽略了那些商业产品中定义的术语。尽管简要定义并不能提供术语的全面解释，但足以使对术语感到困惑的读者有更清晰的了解。索引中给了书中对相关术语有更深入解释的页数。

10Base-T Ethernet　以太局域网的一种特定的布线方案。"T"表示双绞线，是将计算机连接至网络的一种线路类型。

A-to-D converter　模/数转换器。Analog-to-Digital converter 的缩写。

ACK　确认。acknowledgement 的缩写。

acknowledgement　ACK。接收方发送的，表示信息已成功接收的一条应答。在包交换网络中，"确认"是数据到达后所返回的一个数据包。因此，当两台计算机交换信息时，数据数据包和确认数据包沿两个相反的方向传送。

active Web page　动态网页。一种含计算机程序的网页。动态网页在下载后可继续显示更新。

address　地址。分配给一台计算机的一个数值，与给一个家庭分配的一个电话号码相似。当一个数据数据包和确认数据包在两台计算机传输时，该数据包中包含这两台计算机的地址。

ADSL　非对称数字用户线（*Asymmetric Digital Subscriber Line*）。数字用户线（DSL）技术的一种特殊类型。电话公司用它来向家庭和商业用户提供高速 Internet 连接。ADSL 和普通电话并不冲突，也就是说，在数据传送的同时可以正常使用电话。

Advanced Research Projects Agency　（美国国防部）高级研究计划署（ARPA），是为 ARPANET 及 Internet 投资的一个美国政府部门。ARPA 曾在一段时间内被称为 DARPA。

afaik　据我所知（*as far as I know*）。被用于诸如电子邮件和即时消息等电子通信中。

American Standard Code for Information Interchange　美国信息交换标准代码（ASCII）。Internet 使用的一种字符代码。ASCII 给每个字符、数字和标点符号分配一个惟一的二进制数字序列。文本数据在 Internet 传输时，常以 ASCII 形式表示。

analog　模拟。一种信息的表示形式。在模拟表示中，物质或信号的量和所表示的信息成一定比例。

Analog-to-Digital converter　模/数转换器（*A-to-D converter*）。一种将模拟电信号转换成数字序列的电子部件。

anonymous FTP　匿名 FTP。使用这种特殊的匿名登录，可通过 FTP 服务访问到各种公用文件。

applet　小应用程序。使用 Java 技术编写的动态网页。小应用程序是能在浏览器屏幕上支持平滑动画显示的计算机程序。

ARPA　（美国国防部）高级研究计划署。*Advanced Research Projects Agency* 的缩写。

ARPANET 由 ARPA 投资兴建的一种早期的广域网，该网络在 1969～1989 年期间一直是早期联网研究的基础。在 Internet 开发期间成为中心主干网。

ASCII 美国信息交换标准代码。*American Standard Code For Information Interchange* 的缩写。大多数文本使用这种表示方式。

Asymmetric Digital Subscriber Line 参见 ADSL。

Asynchronous Transfer Mode 异步传输模式（ATM）。这是一种特殊网络技术的名称，设计该技术的目的是使网络高速运行。ATM 已经不再流行了，不过它仍在 DSL 调制解调器中被使用。

ATM 异步传输模式。*Asynchronous Transfer Mode* 的缩写。

attachment 附件。MIME（参见 *Multi Purpose Mail Extensions*）格式的电子邮件消息中被一同传送的一项。通常，附件包含一个非文本的文件，例如音频剪辑或照片。

audio clip 音频剪辑。数字化的一段（一般较短）音频。音频剪辑一般存储在文件中，能被包含在电子邮件消息中通过 Internet 传送。

audio teleconference 音频电信会议。一项使一组用户能在 Internet 上交换音频信息的服务，类似于电话会议。每个参与者的计算机必须配备一个麦克风和扬声器（或耳机）。

automated search service 自动搜索服务。任务无须用户做出决策或选择菜单项即可进行信息定位的服务。典型的自动搜索服务能查找整个文件。*Google* 和 *Yahoo!* 都属于这类自动搜索服务。

backbone network 骨干网络。指由点路由器接入的中心网络。在 Internet 上，骨干网使用广域网技术。个别公司也将中心网络称为骨干网，这些骨干网可能是局域网。

bandwidth 带宽。用于表示网络容量的大小，通常以“位每秒”作为单位。在网络系统中，音频和视频传输需要的带宽较高，而电子邮件或其他服务所需要的带宽较低。也见 *broadband*。

baud 波特。从字面上看，波特是指信号在传输线路上每秒钟改变的次数。波特率常用于描述拨号电话连接的速度。

best-effort delivery 尽力而为传送。指计算机能在发生拥挤时丢弃一些数据包，因为 TCP/IP 软件可检测到丢失的数据包并重新发送它们，所以该软件能使用尽力而为传送网络。

binary 二进制。指任何使用二值的数字系统。计算机内的所有运算都使用二进制。

binary digit 二进制数字（*bit*）。0 或 1。Internet 使用二进制数字表示所有信息，如音频、视频和文本。

binary file 二进制文件。指非文本文件，如数字照片或音乐。（参见 *text file*）

bit 比特。binary digit 的缩写。

bits per second 位每秒（bps）。数据传输速率，用于衡量网络的容量（参见 *bandwidth*）。现代网络以百万位每秒（Mbps）运行。

bookmark 书签。浏览器中的一个工具，用于记录一个特定页面的位置，以便于以后返回到该页。书签集也称作收藏夹（*favorites*）。

bps 位每秒。*bits per second* 的缩写。

broadband 宽带。用于描述高速运行的计算机网络。DSL 和线缆调制解调器技术用来向居民和商业用户提供宽带 Internet 访问。

broadcast 广播。一种数据包传送机制，可将一个给定数据包的副本传送给连接在网络上的所有计算机。与单播（unicast）和多播（multicast）比较而言。

browser 浏览器。使用户能查看到万维网上的超媒体文档的计算机程序。Mozilla 和 Inter-

net Explorer 这两种浏览器可以分别从 Netscape 通信公司和 Microsoft 公司获得。

browsing 浏览。通过重复扫描和选择来查看信息。Internet 的浏览服务首先显示一页信息，在用户读完信息并选择了一项后，浏览服务按照所选项目返回新的信息。

btw 顺便。电子通信（尤其是即时消息）中使用的 *by the way* 的缩写。

bulletin board service 公告板服务。一项允许一个人发布一条消息供其他人阅读的服务。每个公告板一般包括关于某个主题的讨论。

cable modem 电缆调制解调器。是一种使得 Internet 服务与有线电视使用同一同轴电缆的设备。电缆调制解调器不与正常的电视接收器相冲突，两者可以同时工作。

carrier 载波。是一种稳定的电信号或特定频率的连续正弦波信号。调制解调器利用它为通信线路或电话连接上传输的信息编码。当计算机使用拨号调制解调器在电话连接上进行通信时，载波是在音频范围内的连续正弦波信号。

Cc 副本（*Carbon copy*）。电子邮件头部中的一行，用于指定附带的接收者。

Central Processing Unit 中央处理单元（CPU）。在计算机中执行所有算术和逻辑运算的电子部件。

CGI 通用网关接口。参见 *Common Gateway Interface*。

chat room 聊天室。一项允许用户加入讨论的服务。当某一参加者输入信息，聊天室中的所有其他参加者都能接收到一个副本。与即时消息（*Instant Messaging*）比较。

checksum 校验和。用于检测数据在两台机器间传输时是否出错的一个小整数值。如 TCP 之类的协议软件计算校验和，并把该值添加到要发送的数据包中。接收端的协议软件重新计算校验和，并把该值和发送来的值比较，以验证数据包内容的正确性。

chip 芯片。集成电路（*Integrated Circuit*）的一种非正式术语。

client 客户。一个通过 Internet 和一个远程服务器通信的程序。通常每项 Internet 服务都需要一个独立的客户程序。但是，浏览器包含多种客户程序，这意味着客户方的浏览器可以获得多种 Internet 服务。

client-server computing 客户机/服务器计算。指两个程序在通过网络进行通信时的交互。某一站点的程序向另一站点的程序发送一个请求，然后等待应答。发出请求的程序称为客户程序；满足该请求的程序称为服务程序。

Common Gateway Interface 通用网关接口（CGI）。每当用户请求一个网页时，该技术便使用一个计算机程序去生成该网页。使用 CGI 技术所构成的网页是动态的；与静态网页不同，这些动态网页在请求到达后才存储在服务器的磁盘上。

computer conference 计算机会议。一组人分别利用计算机进行通信，计算机会议服务可提供文本的、音频的或者视频的通信。也可参见 *bulletin board service*。

computer network 计算机网络。一种用于计算机通信的硬件机制。按硬件容量的不同，网络可分为两类：广域网和局域网。

congestion 阻塞。计算机网络中，当过多数据包同时到达网络中的某一节点时出现的问题。与在普通的高速公路上相同，阻塞意味着高延迟。

connection 连接。用来描述两个应用程序首先同意通信，然后交换数据的计算机交互过程。TCP 软件使用连接式的交互，意味着大多数的 Internet 服务使用连接，与无连接（*connectionless*）相对。

connectionless 无连接。用来描述一个应用程序不必等另一个应用程序同意通信就向它发送数据或请求。网际协议，即 IP，就是使用非连接技术。

cookie 用于确认万维网用户身份的少量数据（通常为少于 50 个字符的字符串）。当某用户访问万维网时，服务器要求器浏览器存储一个 cookie 以便于该用户下次访问时确认其身份。

CPU 中央处理单元。*Central Processing Unit* 的缩写。

CSNET 计算机科学网（*Computer Science NETwork*）。一个早期的网络，它提供各学院和各大学内计算机科学系之间的电子邮件和 Internet 的连接。CSNET 最初是由美国国家科学基金会（NSF）投资兴建的，后来开始自给自足。

D-to-A converter 数/模转换器。*Digital to Analog converter* 的缩写。

DARPA 美国国防部高级研究计划署。*Defense Advanced Research Projects Agency* 的缩写。原名为 ARPA。

datagram 数据报。与 IP 数据报同义。

decryption 解密。对加密过以保证其机密性的消息进行解码的过程。在消息通过 Internet 的传输到达接收方后，由接收方对其进行解密。

Defense Advanced Research Projects Agency 美国国防部高级研究计划署（DARPA）。原名为 *Advanced Research Projects Agency*。

demodulation 解调。从传输线路或电话连接上传来的经过调制的信号提取出信息的过程。解调通常在调制解调器中进行，也可参见 *modulation* 和 *carrier*。

demodulator 解调器。调制解调器中用于给输入的信号解码，以提取数据的电子设备。参见 modulator。

destination address 目的地址。一个数据包中用于指定数据包发往目的计算机的数值。在 Internet 上传送的数据包，其目的地址时目的计算机的 IP 地址。

dial-up access 拨号连接。用普通的电话拨号连接 Internet 的技术。拨号连接是最便宜的也是最慢的一种连接方式。

digital 数字。指那些使用数字来表示信息的技术。从本质上说，计算机是数字式的，因为它使用数字来表示按键的输入、图像、文本、声音和视频。

digital camera 数码相机。是一种相机，它将图像数字化，并储存成更适合于计算机而不是胶卷的形式。

digital cash 数字现金。一种使 Internet 上的顾客不必为小额支付逐个填写信用卡信息的技术。许多提供数字现金服务的公司都失败了。

digital certificate 数字证书。由可信的权威机构发送的，用于安全性事务的消息。数字证书包含一个公钥的副本。

digital library 数字图书馆。数字形式存储的信息的集合。数字图书馆包括文件、图像、声音、书本、视频以及从当前事件（如从气象卫星传来的连续的图像）中收集的信息。

digital signature 数字签名。用于认证某文件作者的加密的消息。数字签名不能被伪造。

Digital Subscriber Line 参见 *DSL*。

Digital-to-Analog converter 数/模转换器（*D-to-A converter*）。将数字序列转换成模拟信号的电子设备。数/模转换器可将光盘上的数字转化为声音。

digitized 数字化。转化为数字序列的信息。扫描过的图像为已被数字化。

directory 目录。文件和其他目录的集合。一些计算机将目录称为文件夹（*folder*）。

distributed computing 分布式计算。用来描述涉及一台以上计算机的计算。从广义上说，每项 Internet 服务都使用分布式计算。

DNS 域名系统。*Domain Name System* 的缩写。

domain name 域名。给 Internet 上的计算机分配的名字。一台计算机的名字常包含多个字符串，各串之间由小数点隔开（例如，computer1.company.com）。域名经常以 .com 或 .edu 结束。

Domain Name System 域名系统（DNS）。用于查找计算机名字和 IP 地址的 Internet 服务。

dotted decimal 带点十进制。该表示法用于表示 IP 地址。带点十进制用四组较小的十进制整数来表示地址，每组数字间用小数点隔开。在计算机内部，IP 地址以二进制形式存储。带点十进制表示法只是为了便于人们的输入和阅读。

DSL 数字用户线（*Digital Subscriber Line*）。指任何使用电话服务线路提供高速 Internet 连接的技术。DSL 和普通电话并不冲突，在数据传送的同时可以正常使用电话。参见 *ADSL* 和 *xDSL*。

dynamic content 动态内容。指任何在浏览器请求下才由服务器生成的网页。只有包含动态内容的网页才能提供即时信息（例如，股票报价）。

e-commerce 电子商务。*electronic commerce* 的缩写。

email 电子邮件。*electronic mail* 的缩写。

email address 电子邮件地址。每个用户分配到一个电子邮箱地址。发送电子邮件时，用户必须输入接收人的电子邮件地址。Internet 上的电子邮件地址形式通常为 *user@computer*。

email alias 电子邮件别名。是电子邮件地址的简记形式，可使用户在发送电子邮件时不必记住或键入冗长的电子邮件地址。大多数电子邮件软件允许用户定义多个别名。

electronic bulletin board service 电子公告板服务。与 *bulletin board service* 同义。

electronic commerce 电子商务（*e-commerce*）。泛指任何在 Internet 上进行的交易。消费者购买是一种流行的电子商务形式。

electronic mail 电子邮件（email）。一种允许一个人向另一个人、某个团体、或某个计算机程序发送一条消息的服务。电子邮件软件也允许某人应答一条消息。

encryption 加密。对消息进行编码以保证其机密性的过程。一条消息在经过 Internet 传输后，由接收方对它进行解密。

Ethernet 以太网。由 Xerox 公司发明的一种流行的局域网技术。现代以太网技术中，几台计算机通过连接到一个小型设备来上网。计算机与 DSL 或线缆调制解调器之间的连接就属于以太网。

exponential growth 指数增长。用于描述 Internet 增长的数学术语。Internet 的规模大约每年增长一倍。

FAQ 常见问题（*Frequently Asked Questions*）。一个包括关于某个特定主题、技术或 Internet 服务方面常疑问和解答的文件。许多邮寄表和新闻组中都有 FAQ 文件，以帮助初学者理解讨论的目的，避免一些基本问题的不断出现。

fiber 光纤。*optical fiber* 的非正式术语。

file server 文件服务器。是一种计算机程序，可提供对该机器上文件的访问。该术语经常指运行文件服务器的计算机。

File Transfer Protocol 文件传输协议（*FTP*）。一种用于在两台计算机间传输文件拷贝的 Internet 服务。FTP 用来下载文件。

finger 查找器。是一种 Internet 服务，用来判断当前哪些用户登录远入某台计算机，或查找关于某用户的详细情况。出于安全性的考虑，Internet 上几乎没有站点提供 finger 服务。

firewall 防火墙。在某站点与 Internet 之间存在的一种安全机制，以保护该站点的计算机和

网络不受攻击。小型的个人防火墙可以保护居民用户；防火墙软件也可以保护个人计算机。

flame 火焰。是电子通信方面使用的一个俚语，表示"令人感动或激动的注释"，常常是为了答复另一条消息而写的。该词有时用作口语，意思为"写一条令人感动的消息"。

flow control 流量控制。控制一台计算机向另一台计算机发送数据的速率。在 Internet 上 TCP 软件提供流量控制，使高速计算机能够与低速计算机进行通信。

folder 文件夹。与目录（*directory*）同义。

forms 表单。一种允许用户和网页进行交互的技术。浏览器显示一个对话框和一些按钮，用户可将文本输入对话框，并且选择按钮。

frames 框架。一种用于万维网的技术，把一个网页分成多个区（即窗口），并允许独立改变各区的内容。各公司常将一条广告放入一个框架中，同时允许用户选择其他框架中的内容。

Frequently Asked Questions 参见 *FAQ*。

FTP 文件传输协议。*File Transfer Protocol* 的缩写。

FYA 让你高兴。电子通信中使用的 *For Your Amusement* 的缩写。

FYI 让你知道。电子通信中使用的 *For Your Information* 的缩写。

GIF 交互式图形格式（*Graphics Interchange Format*）。表示数字化图像的一种格式，在万维网上流行。也可参见 JPEG。

gopher 基于菜单驱动的 Internet 信息查询工具。一种早期的 Internet 浏览服务的名称。它将所有信息组织成分层菜单的形式。gopher 将菜单显示在屏幕上来让用户挑选菜单项。选项指向一个信息文件或另一个菜单。

GUI 图形用户界面（*Graphical User Interface*）。一种早期给用户图片或图标，而非文字的接口。

hacker 黑客。起初指拥有罕见才能的程序员，现在指侵入别人计算机的人。各公司使用各种安全技术，例如防火墙，来防治黑客攻击他们的计算机。

homepage 主页。通过万维网可访问到的信息页。是个人或公司网络的主页面。主页里通常包括其他页的引用。

hop count 跳步数。用于衡量包交换网络的距离。如果一个数据包在从源端到达目的端的路径上必须经过 *N* 个路由器，则称这两端间存在 *N* 个跳步。

host 主机。即用户的计算机。从技术角度说，与 Internet 连接的计算机可分为两类：主机和路由器。

hostname 主机名。分配给计算机的名字。参见域名（*domain name*）。

hot list 热单。该术语最初表示一列书签或收藏夹，现在基本不用。

HTML 超文本标记语言（*HyperText Markup Language*）。一种计算机语言，用来指定万维网中超媒体文档（例如，主页）的内容和格式。用户在浏览时不可能看到网页的 HTML 形式，因为浏览器自动解释 HTML，并显示其结果。

HTTP 超文本传输协议（*HyperText Transfer Protocol*）。用于访问万维网文件的协议。用户在统一资源定位符（URL）中可能会遇到 *http* 字符串。

hub 集线器。用于连接几台计算机，作为局域网（通常指以太网）核心的电子设备。

hypermedia 超媒体。一种信息存储系统，其中包含信息能包含图像、声音、视频和其他信息页的嵌入式引用。当用户选择其中的一项后，超媒体系统将调出相关的引用。参见万维网（*World Wide Web*）。

hypertext 超文本。存储文本信息页的系统，其中每一页都包含对其他信息页的嵌入式

引用。

HyperText Markup Language 参见 *HTML*。

HyperText Transfer Protocol 参见 *HTTP*。

IAB Internet 架构委员会，*Internet Architecture Board* 的缩写。

IETF Internet 工程任务组，*Internet Engineering Task Force* 的缩写。

IM 即时消息，*Instant Messaging* 的缩写。

imho 愚见，电子通信，如电子邮件、即使消息中使用的 *in my humble opinion* 的缩写。

information browsing service 信息浏览服务，一种允许用户通过重复地扫描和选择以浏览信息的服务，该服务首先显示包括一个或多个选项的信息页；在用户阅读信息并选择一项后，该服务执行用户选择的引用，并返回新信息页。

information superhighway 信息高速公路，新闻界使用它来指美国国家信息基础设施。Internet 是信息基础设施的主要部分。

infrastructure 基础设施，社会的基础服务或工具，如运输和供水系统、运输设施、电话系统等。

Instant Messaging 即时消息，一项允许两个用户连接并交换简短文本消息的 Internet 服务。无论一方键入什么，都能显示在另外一方的屏幕上，反之亦然。即时消息服务也已得到扩展，包括计算机会议（一组用户），音频和视频，简称为 IM。

integrated circuit 集成电路，一个包含许多晶体管的小而复杂的电子器件。例如，计算机内的中央处理单元 CPU，通常是在单个集成电路板上制造的，非正式地，集成电路称为芯片（chip）。

International Telecommunication Union 国际电信联盟（*ITU*），电话设备的互联建立标准的国际组织。过去，许多负责在欧洲建造网络的组织都服从 ITU 的建议标准。（ITU 原来称为 CCITT）[9]

Internet 因特网，一个统一的大型网络，它由使用 TCP / IP 协议族及其功能的数目庞大的网络及路由器构成。因特网已经遍及政府、商业及全世界的教育组织。

Internet address Internet 地址，用于标识计算机的一串数字，接入 Internet 的每台计算机都被分配一个惟一的地址，当某个软件要发送一则消息时，它靠 Internet 地址来识别接收者。Internet 地址也被称作 IP 地址。

Internet Architecture Board Internet 架构委员会，缩写 *IAB*，对 TCP / IP 及 Internet 制定策略和标准的组织。

Internet Engineering Task Force Internet 工程任务组，缩写 *IETF*，负责设计、测试 TCP / IP 和 Internet 领域新技术的国际组织，IETF 是 IAB 的组成部分。

Internet Fax Internet 传真，一种通过 Internet 发送的传真，该传真不同于传统的电话传真方式。Internet 传真在使用时需要接入 Internet。

Internet Protocol 参见 *IP*。

Internet Relay Chat Internet 中继聊天，一种允许用户使用键盘进行交互的通信服务，每组用户建立一条信道，然后发送消息。每个处于活动状态的参与者都可以接收到发给该通道的所有信息。

Internet Service Provider 参见 *ISP*。

---

[9] International Telephone and Telegraph Consultative Committee，国际电话与电报顾问委员会。

Internet Society　Internet 团体。鼓励研究 Internet 的非盈利性组织。

internetworking　Internet 技术，指设计、建造、测试和使用 Internet 系统的各项技术。

IP　网际协议，是 *Internet Protocol* 的缩写。从字面上讲，IP 指的是计算机通信在 Internet 上所采用的数据数据包格式。实际上，IP 常指代计算机在 Internet 上通信时所必须运行的 IP 软件。

IP address　IP 地址。与 *Internet address* 同义。

IP datagram　IP 数据报。指的是在 Internet 上传送的数据报文数据包，每个 IP 数据报都包含发送它的计算机的地址、目的计算机的地址以及发送的数据。

IPv6　网际协议第 6 版。是继 IPv4 之后的最新版本协议。IPv6 由 IETF 开发，但之后 10 多年来始终未能产业化。

IRC　Internet 中继聊天，是 *Internet Relay Chat* 的缩写。

ISP　Internet 服务供应商，是提供与 Internet 连接的公司。除了电话公司和有线电视公司之外，还有许多规模较小的私营 ISP 在较小区域（例如，在一座城市）内提供服务。

ITU　国际电信联盟，是 *International Telecommunication Union* 的缩写。

Java　爪哇。是一种由 Sun Microsystems 公司开发的，用来制作动态 Web 网页的程序设计语言。参见 applet。

JavaScript　Java 脚本语言，是一种制作动态 Web 网页的程序设计语言。与 Java 相比，JavaScript 制作网页更加简单，功能也比较单一。

JPEG　Joint Photographic Experts Group 的缩写，一种数字图像压缩格式。JPEG 在万维网上十分流行。参见 *GIF*。

Kbps　千位每秒。用来衡量数据传输率，其值等于 1000bps。参见 bps、Mbps、bandwidth 以及 baud。

key　密钥。通过加密数据确保其安全的字符串。人们可以把密钥看作密码。典型的密钥长度在 56 位到 256 位之间，密钥长度越长安全性越高。

LAN　局域网，是 *Local Area Network* 的缩写。

last mile problem　最后一英里问题。指的是居民住宅区如何接入 Internet 的问题。参见 DSL 和 cable modem。

Linux　UNIX 操作系统的运用版本实现。参见 UNIX。

LISTSERV　电子邮件列表服务器，是 *electronic mailing LIST SERVer* 的缩写。一个用来维护电子邮件地址列表的程序。用户可请求 LISTSERV 将他的 email 地址加入表中或从表中删除。

Local Area Network　局域网，缩写 LAN。一种短距离实现计算机互联的网络技术。例如，在一座建筑物内。与广域网（WAN）相对。

login　登录。通过输入账户标识符和口令以获取访问计算机的过程。

long-haul network　远程网。与 *Wide Area Network* 同义。

mail alias　email 别名。与 *email alias* 同义。

mailbox　邮箱。指的是一块存储区域（通常在硬盘上），它保存着收到的所有邮件。每个邮箱都有一个惟一的地址；要接收电子邮件用户必须拥有自己的邮箱。

mailbox address　邮箱地址。与 *email address* 同义。

mailing list　邮件列表。系列电子邮件地址，该地址中包括一系接收列表，作为一种广播信息的手段，邮件列表已经非常受欢迎。

Mbps　百万位每秒，用来衡量数据传输率，其值等于 1 000 000bps。参见 *bps*、*Kbps*、

bandwidth 以及 *baud*。

menu 菜单。用户可以选择的一系选项。一些 Internet 服务允许用户通过选择一系列菜单来浏览信息。

Micropayment 小额付款机制，指网络中的小额交易方式。

MIME 多用途 Internet 邮件扩展，是 *Multipurpose Internet Mail Extensions* 的缩写。

modem 调制解调器。一种用来在模拟传输线路上远程传输数字信息的设备，这条传输线路可能是一条长导线，也可能是一条通过拨号电话系统连接的线路。调制解调器是成对使用的，即在连接线路的每一端都连接一个调制解调器。调制解调器由调制器（用来发送数字信息）和解调器（用来接收数字信息）组成。

moderated 受限新闻组。一类邮件列表或网络新闻讨论组，其中每个成员提交的内容都必须发送给一名负责检查和编辑这些内容的人员，最后把这些内容转发给接收者。

modulation 调制。是一种使用调制解调器将数字信息编码成电信号，以便在线路上进行传输的技术。调制技术也可在电话连接线上使用。参见 *demodulation* 和 *carrier*。

modulator 调制器。调制解调器中用来对传输的数据进行编码的电子设备。参见 *demodulator*。

MP3 一种音频编码格式。MP3 广泛用于存储音乐。

MUD 泥巴游戏。是 *Multi-User Dungeon* 的缩写。一种互联网游戏的名称。

multicast 多播。该技术用于将一个数据包的多个拷贝发送到预先指定的其他多个计算机上。Internet 音频和视频服务，用多播技术从单个数据源发送数据包到 Internet 上的许多计算机上，或者允许一组用户进行音频或视频电话会议。试与 *unicast* 和 *broadcast* 进行对比。

multimedia 多媒体。指能显示文本、图形、图像、视频和声音的设备。计算机中需要有特殊的硬件来控制多媒体的输出。

Multipurpose Internet Mail Extensions 多功能 Internet 邮件扩展，缩写 MIME。尽管标准的电子邮件只包含 ASCII 文本，但 MIME 扩展允许电子邮件中包含诸如图像和声音之类的非 ASCII 文件。非 ASCII 文件常常以附件（*attachment*）的形式作为电子邮件的一部分，附件经过 MIME 编码后发送出去。

navigating the Internet 漫游 Internet。该短语的意思是：使用 Internet 服务来浏览各种信息。

Netiquette 是一张关于在使用 Internet 时应如何操作的各种建议的表格，其中有许多都是常识。

netnews 网络新闻。是 *network news* 的缩写。

Netscape 一家经营 Internet 产品的公司。Netscape 出品的 Web 浏览器 Mozilla 已经广为人知。该浏览器可免费使用。

network news 网络新闻。Internet 公告板服务的名称。

network of networks 网中网。用于描述 Internet，它是由许多网络通过路由器互联构成的。

network printer 网络打印机。直接连接在网络上的打印机，网络上的任何计算机均可使用它。网络打印机中含有一个小型接处理器，以处理通信细节。

network provider 网络提供商。参见 ISP。

news article 新闻文章。显示在网络新闻服务公告板上的消息。新闻文章的形式和电子邮件备忘录相同。

newsgroup 新闻组。是一个网络新闻服务公告板。一个用户订购多个新闻组，每个新闻组中包含许多关于某个主题的文章。

NFS 网络文件系统，是 *Network File System* 的缩写，一项允许多个相互协作的计算机互相访问对方的文件系统（就像它们位于同一地方一样）的服务。NFS 和 FTP 的主要区别是 NFS 按需访问文件中的某些部分，而不是拷贝整个文件。

NSF 国家科学基金会，是 *National Science Foundation* 的缩写，美国一个政府部门，它为 Internet 投资兴建了一个 WAN，并帮助科学家建立与 Internet 的连接。NSF 也为一些从事网络领域研究的单个研究人员，以及跨越多个研究所的研究计划提供资金。

NSFNET 国家科学基金会网络，是 *National Science Foundation NETwork* 的缩写。它指的是在 20 世纪 90 年代中期形成的美国国内 Internet 主干广域网。

open system 开放式系统。一种非专用的技术或系统，任何厂商都能使用开放系统的标准来开发产品和服务。Internet 及其技术是开放的。

optical fiber 光纤，是一种细且有弹性的玻璃纤维，它用光波来传输信息。光纤能够比电缆跨越更远的距离。*optical fiber* 常被缩写成 *fiber*。

packet 数据包，数据包是交换网络上传输的数据单元的统称。Internet 的数据包称为 IP 数据报。

packet switching 包交换技术，是一种在计算机网络上使用的技术，该技术要求计算机在发送信息之前先将信息分解为一组数据包。Internet 和大多数计算机网络一样，也使用包交换技术。

password 口令，用户为了进入分时系统或得到访问 FTP 服务授权而必须输入的秘密代码。通常计算机并不显示用户输入的口令。

peer-to-peer networking 对等网络，所有计算机都是平等的网络系统。在一台计算机上运行的程序能和在另一台计算机上运行的程序相互联系。Internet 可看作一种对等技术。

personalized search results 个性化的搜索结果，从被发出请求的用户个人化的搜索引擎得到的结果。个人化依赖于搜索和 Web 访问的历史信息。

PGP 是 *Pretty Good Privacy* 的缩写，是一种用于保护邮件私密性的加密机制。

ping 数据包网络测试程序，是 *Packet InterNet Groper* 的缩写。TCP/IP 网络中测试是否能和其他计算机联系上的程序的名称。ping 先向计算机发送一组数据包，然后等待应答。人们常会说："请 ping 一下计算机，看它是否是活动的。"

plugin 插件，使用该技术，浏览器能动态地装入附加软件，以便解释新的或其他的数据格式。浏览器只有使用插件技术，才能播放 Web 网页上的声音记录。

point-and-click interface 点击界面，其特点是使用鼠标而不是键盘和计算机进行交互。用户通过移动鼠标来确定光标的位置，然后通过点击鼠标上的按钮来选择光标下的菜单项。

POP 是 *Post Office Protocol* 的缩写。

Post Office Protocol 缩写为 *POP*。它是一种访问邮箱、下载电子邮件的应用协议。POP 常被提供拨号访问服务的 ISP 使用。

Post, Telegraph, and Telephone 缩写为 PTT。欧洲各国控制数据网络的政府组织。

postmaster 邮件管理员，根据约定，"邮件管理员"可作为管理一台指定计算机上所有电子邮件的人员的电子邮件别名。人们可以通过向 *postmaster@site* 发送邮件的方式来询问关于站点或者加入或移出邮寄表的规则的问题。

Pretty Good Privacy 参见 *PGP*。

private email 加密电子邮件，使用密钥进行加密，使之成为仅为发送者和接收者能理解的电子邮件。通常在 Internet 上传送的电子邮件消息是没有加密的，这意味着接收电子邮件的人能

够读懂其中的内容。

private key 私钥。为了安全而提供给用户的两个密钥之一。用户必须保密自己的私钥。参见 *public key*。

protocol 协议。两台或多台计算机交换信息时所必须遵守的规则。协议描述了信息传输时所采用的格式，以及计算机对每条信息的应答方式。

PTT 是 *Post, Telegraph, and Telephone* 的缩写。

public files 公共文件。能被 Internet 上任何用户访问的文件。当使用 FTP 服务时，用户可以通过匿名登录来访问公共文件。

public key 公钥。为了安全而提供给用户的两个密钥之一。在加密通信中，用户需要告诉通信对方他的公钥。参见 *private key*。

public key encryption 公钥加密技术。一种安全系统，在该系统中每个用户都被分配一对密钥，其中之一必须保密，而另一个可以被公开发布。参见 *private key* 和 *public key*。

public mailing list 公共邮件列表。一种任何人都可以加入、退出或发送备忘录的电子邮件列表。每名成员都可以看见发送到公用邮件列表的备忘录。

RealAudio 是一种在 Internet 上发送音频的流行技术。许多广播站都使用该技术传送节目。浏览器必须利用插件技术来接收和播放采用实音频编码的声音。

remote desktop 远程桌面。一个允许一台计算机上的用户将显示器、键盘和鼠标连接到另一台远程计算机上的系统。参见 *remote login*。

remote login 远程登录。该服务允许一台计算机上的用户将他的键盘、面向字符的显示器连接到远程计算机，以便运行程序。参见 TELNET。

Request For Comments 缩写为 RFC。是一系列包含 TCP/IP 协议标准及其相关文档的记录。我们可以从 Internet 上获得 RFC 文档。

RFC 是 *Request For Comments* 的缩写。

route 路由。一般来说，路由是指网络传输从源端到目的端所经过的路线。在 TCP/IP 互联网上，每个 IP 数据报的传输路径包括一系列的网络和路由器。

router 路由器。指连接两个或多个网络，并在网络间传递 IP 数据报的一种具有特定功能的专用计算机。每个路由器不断将数据报发送给另一个路由器，直到将数据报传送到目的地为止。

search engine 搜索引擎。应用于自动搜索服务的术语。从技术上说，该术语是指自动搜索服务使用的用于应答用户请求的计算机程序。

search key 查找关键字。指用户给搜索服务提供的一串字符。搜索服务查找包含该字符串的标题或文件。

search tool 搜索工具。指帮助用户查找信息所在位置的程序。特别是，自动搜索工具在搜索时无须和用户交互。

Secure Multipurpose Internet Mail Extensions 参见 *SMIME*。

Secure Socket Layer 参见 *SSL*。

server 服务器。指提供服务的程序。Internet 上的许多计算机都运行服务程序以提供服务。用户调用他的计算机上的客户程序，客户程序将和远程计算机上的服务建立联系。

shopping cart 购物车。应用在万维网上的一种机制，该机制允许顾客从在线货架上挑选多个商品，然后再一起购买。

smiley 在电子邮件中常见的字符序列，表示该邮件中带点幽默。三字符序列 ":-)" 尤其流行，因为它与笑脸的侧视图相似。

SMIME 是 *Secure Multipurpose Internet Mail Extensions* 的缩写。它是一种电子邮件安全机制，用于保护邮件信息的私密性。

source address 源地址。指发送数据的计算机的地址。每个数据包中包含源计算机和目的计算机的地址。

spam 垃圾邮件。常常是一些广告信息，这些垃圾邮件被成批地发送给大量的接收者。

spider 网络爬虫。搜索引擎公司使用的计算机程序，该程序探查整个 Web，然后编辑查到的内容的索引信息。

SSL 是 *Secure Socket Layer* 的缩写。一种由 Netscape 有限公司发明的机制，该机制在服务器与浏览器之间提供安全通信服务。

stack 协议栈。指计算机上若干 TCP/IP 软件。该术语源于软件的内部组织形式。

surfing the Internet 网上神游。该短语的意思是：使用 Internet 服务浏览各种信息。

talk 会话程序。早期的一种即时通信服务，该服务在 UNIX 类作系统中使用。

TCP 传输控制协议，是 *Transmission Control Protocol* 的缩写。

TCP/IP 从字面上讲，它是指定 Internet 上的计算机如何通信的协议名称。另外，该名称也泛指实现该协议的软件。所有使用 Internet 的计算机都需要 TCP/IP 协议。

TCP/IP Internet Protocol Suite TCP/IP 的正式名称。

TELNET Internet 上的远程登录服务。TELNET 允许用户和另一个站点的远程分时系统进行交互，就像该用户的终端直接和远程机器相连一样。

text file 文本文件。由若干行字符组成的文件。Internet 上的文本文件采用 ASCII 字符编码。非文本文件常被称作二进制文件。

textual interface 文本界面。它的典型风格是用键盘与计算机进行交互，用户键入命令，计算机作出响应。试与 *point-and-click interface* 以及 *GUI* 比较。

timesharing computer 分时计算机。允许多个用户同时运行各自程序的计算机系统。大多数计算机都设计成分时系统。

traceroute 路径跟踪。一种允许用户查找数据包通过 Internet 到达目的地所经过的路径的程序。traceroute 每遇到一个路由器便打印出一行。

Transmission Control Protocol 传输控制协议，缩写为 TCP。两个主要的 TCP/IP 协议之一。TCP 负责保证数据以正确的顺序到达目的地。TCP 也常指实现 TCP 标准的软件。

traveling the information superhighway 畅游信息高速公路。这个短语的意思是：使用 Internet 服务浏览各种信息。

Trojan horse 特洛伊木马。一种安全攻击技术，运用该技术，攻击者诱骗公司人员运行一个计算机程序，这个程序悄悄地把此人的信息从本地计算机传送到攻击者指定的目的主机。

unicast 单播。一种常用技术，该技术通过 Internet 从单个源端向单个目的端发送数据包。试与 *broadcast* 以及 *multicast* 相比较。

Uniform Resource Locator 统一资源定位符，缩写为 URL。浏览器用来在万维网上识别特定信息页的简短字符串。有了 URL，浏览器就能够快速地提取、显示指定的页面。

UNIX 由美国电话电报公司贝尔实验室开发的一种特殊的计算机操作系统。Web 服务器广泛采用 UNIX 操作系统。参见 *Linux*。

unsecure 指计算机或网络中的信息没有得到保护，通过有效的保护可以避免非法读取、复制或者篡改。

URL 是 *Uniform Resource Locator* 的缩写。

UUCP　Unix之间的拷贝程序，是 *Unix to Unix Copy Program* 的缩写。该软件是20世纪70年代中期开发的，它允许两台计算机在一条线路上连接（通常是拨号连接），相互拷贝文件。UUCP在 USENET 的早期版本中使用。

video teleconference service　视频电话会议服务。一种允许一组用户在 Internet 上交换视频信息的服务。大多数视频会议包括音频会议设备，每个与会者的计算机必须配备摄像头、麦克风和扬声器（或耳机）。

virtual network　虚拟网络。尽管 Internet 是由许多网络通过路由器互联而形成的，但是通信软件使 Internet 看上去就像一个单一的大型网络。"虚拟网络"这一术语是指表面上单一的、无缝的网络系统。

Virtual Private Network　参见 *VPN*。

VPN　虚拟专用网，是 *Virtual Private Network* 的缩写。一种允许雇员远距离工作的技术。VPN 在雇员计算机和雇主单位网络之间建立一条安全的信道，通过该信道 VPN 给雇员提供远程访问服务，雇员就好像在单位工作一样。

WAN　广域网，是 *Wide Area Network* 的缩写。

Web　参见 *World Wide Web*。

Web authoring tool　网页创作工具，用于创建 Web 页面。网页创作工具通常都有一个图形用户接口。

Web browser　Web 浏览器，参见 *browser*。

Web page　Web 页面（或网页）。万维网上可用的一个信息页面。用户一次只能浏览一页。

Web server　Web 服务器，一个在某网站上持续运行的程序，对来自浏览器的请求进行响应，返回每个请求指定的页面。

Web site　网站，从概念上讲，它是某家公司或个人所有的一系列网页的集合。它是通过一台与 Internet 连接且运行 Web 服务程序的计算机实现的：当请求到来时，服务器负责将请求的页面从网站发送给浏览器。大部分公司都有自己的 Web 网站，通常将安装有服务程序的主机以 *www.company.com* 的形式命名。

Webcasting　Web 广播，一项同时把音频（可能是视频）信息发送给多个用户的服务。每个用户通过 Web 浏览器使用此项服务。

whiteboard service　白板服务，这种服务允许一组用户建立一个会话，以便每个组员都能看见或修改显示出来的信息。显示的信息起初可能是空白的或者可能是一个文档资料。当一个组员通过添加文本或图形修改了显示内容，其他所有组员都能立即看到发生的变化。白板服务通常与音频电话会议服务一起使用。

whois　一种用于检索数据库中某个用户信息的 Internet 服务。*whois* 容易泄露用户信息，从而可能带来产生垃圾邮件，因此许多组织禁用了此项服务。

Wide Area Network　广域网，缩写 WAN。一种能够跨越较大地理范围的计算机网络技术。可与局域网（LAN）相比较。WAN 也称作远程网（long-haul network），它的费用开销通常高于局域网。

Wi-Fi　无线保真（无线局域网标准的简称），是 *Wireless Fidelity* 的缩写。一个营销术语，指的是提供（接近响应）局域网服务的无线网络技术。许多笔记本电脑都配置了 Wi-Fi 接口硬件。

window　窗口。屏幕上一块专属于某个特定应用程序的矩形区域。窗口可以重叠，用户可以将某个窗口拖动到其他窗口的上面。当计算机使用 Internet 服务时，一个窗口可以连接到一台

远程计算机上。

wireless network 无线网络。一种使用无线电波代替线缆传输数据的网络。目前存在大量的无线技术，一部分被应用在同一建筑物中，另一部分可以覆盖相当广阔的区域，例如蜂窝电话系统。

World Wide Web 万维网，缩写为 WWW。一种使用超媒体技术来组织信息的 Internet 服务。每个文档都可以包含对图像、音频和其他文档的嵌入引用，用户通过引用浏览相关信息。

WWW 万维网，是 *World Wide Web* 的缩写。

X Window system X 视窗系统。一种运用在 Internet 上的特殊视窗系统，该系统允许每个窗口连接一台远程计算机。

xDSL x 数据用户线，数据用户线的总称。一个代表所有 DSL 服务的术语，其中，ADSL 最为流行。