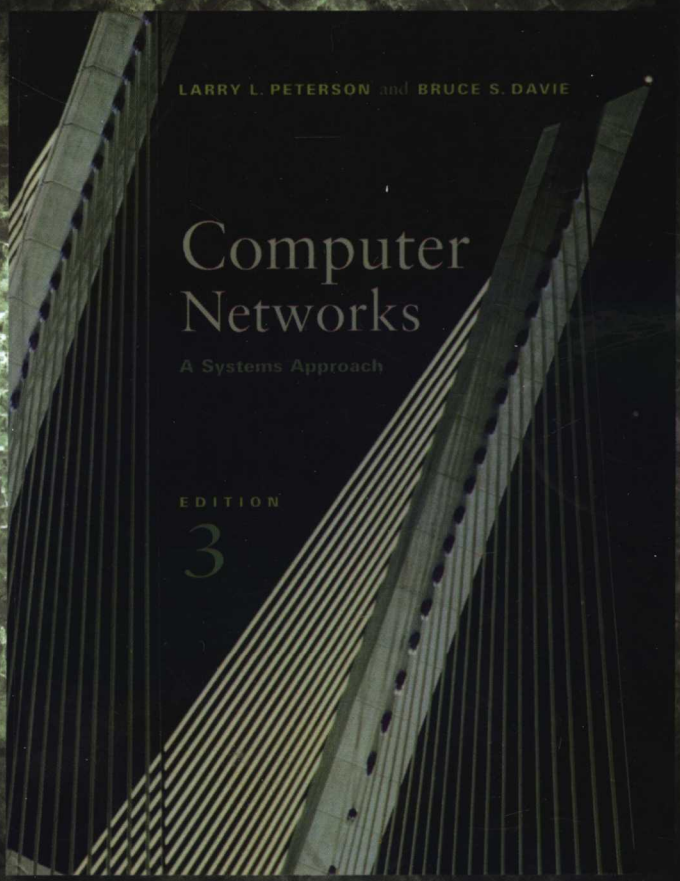
计 算 机 科 学 丛 书

原书第3版

计算机网络
系统方法

（美）Larry L. Peterson　Bruce S. Davie 著　叶新铭　贾波　等译



Computer Networks: A Systems Approach

Third Edition

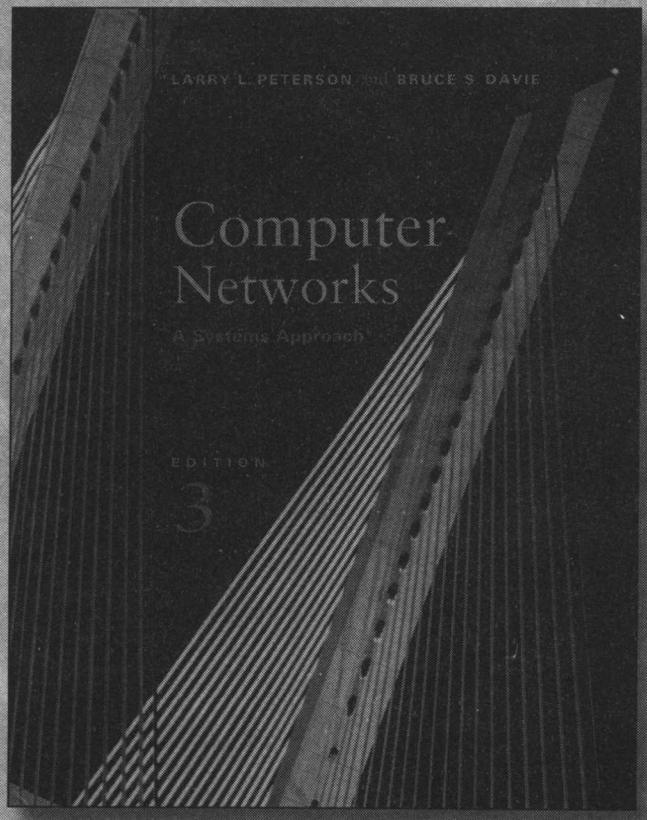
机械工业出版社
China Machine Press

计 算 机 科 学 丛 书

原书第3版

计算机网络 系统方法

(美) Larry L. Peterson Bruce S. Davie 著 叶新铭 贾波 等译



Computer Networks: A Systems Approach

Third Edition

机械工业出版社
China Machine Press

本书介绍计算机网络技术的基本概念和应用，内容详实，论述严谨。本书采用“系统方法”来分析计算机网络，把网络看作一个由相互关联的构造模块组成的系统（反对严格地分层），介绍了很多网络中的新技术，包括对等网络、IPv6、覆盖网、内容分发网络、MPLS与交换、无线与移动技术等，涉及大量的实际应用。本书引入了丰富的因特网实例，说明实际网络的设计，更便于读者理解。每章后的习题有助于读者掌握和复习知识要点。

本书适合作为高等院校计算机及相关专业的本科生和研究生的教材，也适合网络专业人员参考。

Larry L. Peterson, Bruce S. Davie: Computer Networks: A Systems Approach, Third Edition (ISBN 1-55860-832-X).

本书版权登记号：图字：01-2003-8107

图书在版编目（CIP）数据

计算机网络：系统方法（原书第3版）/（美）彼得森（Peterson, L. L.），（美）戴维（Davie, B.S.）著；叶新铭等译. - 北京：机械工业出版社，2005.1

（计算机科学丛书）

书名原文：Computer Networks: A Systems Approach, Third Edition

ISBN 7-111-15514-9

Ⅰ.计… Ⅱ.①彼… ②戴… ③叶… Ⅲ.计算机网络－高等学校－教材 Ⅳ.TP393

中国版本图书馆CIP数据核字（2004）第112254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：李伯民 傅志红

北京昌平奔腾印刷厂印刷·新华书店北京发行所发行

2005年1月第1版第1次印刷

787mm × 1092mm 1/16 · 33印张

印数：0 001-5 000册

定价：49.00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

前　言

当本书第1版在1996年出版时，在因特网上购物还是很新奇的事情，那时如果一个公司用它的域名做广告就被认为是很超前的。而当今社会，因特网商务已进入日常生活中，“.com”股票已经历了一个完整的兴衰循环。从光交换机到无线网络，一大批新兴技术正在成为主流。似乎关于因特网唯一可以预见的东西就是它会不断地变化。

尽管有这么大的变化，我们在第1版中提出的问题对于今天来说仍然是有效的：使因特网得以运行的基本概念和技术是什么？回答是TCP/IP体系结构的大部分功能对于今天仍然适用，这一点正像30年前它的创立者预见的那样。这并不是说因特网的体系结构没什么新鲜的，而是正好相反。一个体系结构30年来不仅幸存下来，而且促进因特网这样快速地增长和变化，了解其中的设计原理正是我们的出发点。正像前两版一样，第3版把因特网的体系结构“何以如此”作为它的基础。

读者对象

我们的目的是把这本书做为广泛的网络课程的教材，供研究生或高年级本科生使用。我们也相信，这本书的核心概念不但对正在进行再培训以便完成网络相关任务的专业人员有吸引力，而且也可以帮助网络从业人员理解每天都要接触的网络协议背后的“为什么”，并且明了网络的整体概念。

根据我们的经验，第一次学习网络的学生和专业人员通常会把网络协议理解成一种从高层传到低层的命令，而他们只要尽量多学一些术语缩写词就可以了。事实上协议是从工程设计原理的应用中开发出来的复杂系统的构件。不仅如此，协议总是根据现实世界的经验不断地被精练、扩展和替换。因此，这本书的目标并不单纯介绍当今使用的协议，更侧重于解释合理的网络设计的基本原理。我们认为把握这些基本原理是应对当今网络领域中的瞬息万变的最好办法。

第3版中的变化

尽管我们关注的是联网的基本原则，但我们使用当今正在运行的因特网中的例子来展示这些原则。因此，我们补充了相当多的新材料，跟踪近期内联网技术的重要进展。我们同时对原有材料做了删除、重新组织和改变侧重点，以反映过去7年发生的变化。

也许自编写第1版以来我们所察觉到的最重要的变化就是现在几乎每一位读者都对诸如万维网和电子邮件这样的网络化应用有一定的了解。因此，我们从第1章开始就加大了对应用的侧重。我们把应用当作学习联网的动机，并得出一组需求，有用的网络只有满足这些需求才能在全球范围内支持当前的和未来的各种应用。然而，我们保留了前两版解决问题的方法，即从主机的互连问题开始，逐层向上讨论，最后对应用层的问题进行详细的考察。我们认为从各种应用及其需求起步对于在本书所覆盖的各主题间建立联系是很重要的。同时，我们感到对于诸如应用层协议和传输层协议这样的高层协议的问题，只有在讲明白主机连接和分组交换这样的

基本问题之后才能很好地理解。

这一版的习题也有重要的改动。我们增加了习题的数量并提高了质量；力求确定那些特别困难或需要较高数学知识水平的习题（这些习题用★标记）；在每一章里还补充了一些在书中可以找到现成答案的习题。

就像我们在第2版中所做的一样，我们附加或增大了对重要的新主题的覆盖面并使其他主题紧跟潮流。这一版中主要的新主题或有实质性改动的主题包括：

- 新增关于多协议标记交换（MPLS）的一节，包括通信量工程和虚拟专用网的内容。
- 新增一个关于覆盖网的一节，包括对等网和内容分发网。
- 增加大量多媒体应用相关协议的内容，比如会话启动协议（SIP）和会话描述协议（SDP）。
- 更新了关于拥塞控制机制的部分，包括TCP的选择应答，基于等式的拥塞控制，以及显式拥塞通知。
- 更新了关于安全的内容，包括分布式拒绝服务（DDoS）攻击。
- 更新了有关无线技术的材料，包括扩展频谱技术和正在兴起的802.11标准。

最后，本书补充了一套全面的实验习题，目的是通过仿真实验来展示关键概念。讨论材料由实验习题所覆盖的各节在页边用图标  标明。下文会讲述本书这一新特点的细节。

系统方法

对于像计算机网络这样动态的和不断变化的领域来说，一本教材能提供的最重要的东西是洞察能力，以便能够区别什么是重要的，什么是不重要的，什么是长久的，什么是表面上的。根据我们致力于网络新技术研究的20年经验，和对本科生和研究生讲授网络最新趋势的课堂反馈，以及把先进的网络产品投放市场的经验，我们提炼出了自己的观点，称之为*系统化方法*，它是本书的精髓。这种系统方法有以下含义：

- 与其接受现成的网络产品作为准则，不如从最基本的原理开始，让你了解当今网络技术的发展过程。这就能让我们解释网络为什么像现在这样设计。根据我们的经验，一旦理解了基本概念，对于遇到的任何新协议，消化和吸收起来都将变得相对容易。
- 虽然材料是围绕传统的网络层次被松散地组织起来，从底层开始沿协议栈向上展开的，但是我们并不采用严格的分层方法。许多主题涉及多层，例如拥塞控制和安全性就是这样，所以我们在传统的分层模型之外讨论它们。简言之，我们相信可以很好的使用分层，但是不必受它的限制。采用端到端的观点常常是更有用的。
- 与其抽象地解释协议如何工作，不如使用当今最重要的协议具体地说明网络是如何工作的，许多协议都是源自TCP/IP因特网的。这就允许我们在讨论中借鉴实际经验。
- 虽然在最底层可以用从计算机销售商购买的硬件建立网络，并且通信服务可从电话公司租用，但是只有软件才可以使网络提供新的服务，并且迅速地适应新的需求。这就是我们为什么强调网络软件是如何实现的理由，而不是只停留在描述所涉及到的抽象算法上。我们还从运行的协议栈中得到展示如何实现某些协议和算法的代码段。
- 网络是由许多组件构成的，而在解决一个具体问题时，基本的方法是忽略一些不重要的因素，而理解所有的组件如何组织在一起，构成一个具有特定功能的网络。所以我们花大量的时间解释网络总体的端到端行为，而不只是个别的组成部分，以便能够理解一个

完整的网络是如何运行的，包括从应用到硬件的所有方面。

- 这种系统化方法包含要进行实验性的性能研究，然后使用从定量分析各种设计选择和指导优化实现这两个方面收集的数据。这种强调经验分析的方法贯穿全书。
- 网络很像其他计算机系统，例如操作系统、处理器体系结构、分布式和并行系统等等。它们都很大并很复杂。为了处理这种复杂性，系统设计者常常提出一组设计原则。我们重点介绍这些贯穿全书的设计原则，并用计算机网络中的例子加以说明。

教学法和特点

第3版我们保留了几个有利于教学的特点:

- *问题*。在每一章的开头，我们描述在网络设计中必须解决的一组问题，由它引出本章探讨的一些主题。
- *相关主题*。本书中，相关主题详细说明要探讨的题目或介绍相关的高级主题。在许多情况下，这些主题与实际中的联网有关。
- *突出的段落*。这些段落归纳了在讨论中得出的重要结论，例如广泛使用的系统设计原则。突出的段落前面带有箭头图标 ▶ 。
- *实际的协议*。虽然本书着重核心概念而不是现成的协议说明，但实际的协议常用来说明大部分重要的思想。因此本书可以用作许多协议的参考源。为了帮助你找到这些协议的描述，每节标题中用括号括起来的是协议名称，指明在那一节定义的协议。例如，5.2节描述可靠的端到端协议的原则，它提供对TCP的详细描述，TCP是这个协议的典型例子。
- *开放问题*。每章的叙述以一个开放讨论的问题结尾，这个问题是研究领域、业界或整个社会正在探讨的课题。我们发现这些讨论能使读者更关心所讨论的网络课题并对其产生浓厚的兴趣。
- *补充读物*。在每一章结尾列有精选的参考书目。这些书目一般包含刚讨论的有关题目的创新性论文。我们竭力推荐高级读者(如研究生)学习这个书目中的文章，以便补充各章所讲的材料。

本书结构和课程使用

本书按以下方式组织:

- 第1章介绍全书使用的核心概念。涉及各种应用，讨论了网络体系结构，并定义通常驱动网络设计的定量性能标准。
- 第2章综述广泛的低层网络技术，从以太网到令牌环再到无线网络。也描述所有链路协议必须解决的许多问题，包括编码，组帧和错误检测。
- 第3章讲述交换网(数据报网与虚电路网)的基本模型，并详细地介绍一种流行的交换技术(ATM)。同时也讨论基于硬件的交换机设计问题。
- 第4章讲述网络互连，并且描述网际协议(IP)的基本原理。这一章讨论的一个中心问题是像因特网这样规模的网络如何对分组进行路由选择。
- 第5章讲述传输层，详细地描述因特网的传输控制协议(TCP)和远程过程调用(RPC)，它们用于建立客户/服务器的应用。
- 第6章讨论拥塞控制和资源分配。这一章的问题贯穿网络层(第3，4章)和传输层(第5章)。

特别注意，这一章描述拥塞控制如何在TCP上工作，并且介绍因特网和ATM为提供服务质量所使用的机制。

- 第7章考虑通过网络发送的数据。这涉及表示格式和数据压缩两方面的问题。压缩的讨论包括解释MPEG视频压缩和MP3音频压缩是如何工作的。
- 第8章讨论网络安全，范围包括加密协议(DES、RSA、MD5)，安全服务的协议(鉴别、数字签名、消息的完整性)以及完全的安全系统(增强型加密邮件、IPSEC)的讨论。这一章也讨论像防火墙这样的实用问题。
- 第9章描述网络应用的典型实例和它们使用的协议，包括像电子邮件和万维网这样的传统应用，和像IP电话和视频流这样的多媒体应用，以及像对等文件共享和内容分发网络这样的覆盖网络。

对本科生的课程，可能需要追加课时帮助学生理解第1章的导论材料，而放弃第6~8章的高级主题。然后在第9章转到网络应用的通常主题上。相反，研究生的指导教师可用一两次课讲完第1章的内容，让学生自己更仔细地研究材料，以腾出更多的时间深入讲授最后四章的内容。研究生和本科生都要完成中间四章(第2~5章)的核心材料。但本科生可有选择地跳过那些更深入的章节(如2.2，2.9，3.4和4.4节).

对于自学本书的读者，我们相信所选的主题涵盖了计算机网络的核心内容，因此建议从前到后顺序阅读。另外我们提供了详细的参考文献目录，帮助读者进一步确定感兴趣领域的补充材料。我们还提供了选题解答。

本书采取独特的方法来讨论拥塞控制，即把有关拥塞控制和资源分配的所有专题集中到第6章。这样做是因为拥塞控制问题不能在任何一层单独解决，同时我们希望读者同时能够考虑各种设计选择(这和我们的观点是一致的，即严格的层次性常常模糊了重要的设计概念)。然而，对拥塞控制的更传统的处理方式是可能的，即在学习第3章时参考6.2节的内容，以及在学习第5章参考6.3节的内容。

习题

在第2版和第3版中都对习题做了重大修改。在第2版我们增添了许多习题，并以课堂测验为基础，大大提高了习题的质量。在这一版我们又增加了少量习题，但做出另外两个重要改动:

- 对那些我们认为很有挑战性或需要本书以外知识（例如概率专门知识）的习题，我们加上★标记以表明它们具有更高层次的难度。
- 在每一章我们都附加了一些范例习题并在书后“选题解答”中给出答案。这些习题用✓标记，目的是为解决书中的其他习题提供一些帮助。

现有的习题分为如下几类:

- 分析性的习题，要求学生做简单的代数计算，展示他们对基本关系的理解。
- 设计问题，要求学生提出和评价各种情况下的协议。
- 动手的习题，要求学生写少量代码行去测试一个想法或使用现成的网络工具进行实验。
- 文献研究问题，能够让学生更深入了解某个特别的问题。

补充材料和在线资源

其他辅助材料，可以在Morgan Kaufmann 出版公司的网站http://www.mkp.com上找到

（搜索（Computer Networks）。

致谢

如果没有许多朋友的帮助本书是不可能问世的。我们非常感谢所有为改进本书做出贡献的人。然而，在致谢之前要提到的是，我们已经尽力改正审阅人指出的错误以及尽量准确地描述同事们给我们解释的协议和机制。如果还有什么错误，那就是我们的责任。如果你发现任何错误，请发电子邮件给我们的出版商Morgan Kaufmann，地址是netbugs@mkp.com，我们将在本书再次印刷时改正它们。

首先，我们衷心感谢审阅过全部或部分手稿的人。除了那些审阅过前两版的人，我们要感谢Carl Emberger, Isaac Ghansah和Bobby Bhattacharjee对全书的审阅。还要感谢Peter Druschel, Limin Wang, Aki Nakao, Dave Oran, George Swallow, Peter Lei和Michael Ramalho 对一些章节的审阅。我们也要感谢所有提供反馈意见和信息来帮助我们决定如何写第3版的人，他们是Chedley Aouriri, Peter Steenkiste, Esther A. Hughes, Ping-Tsai Chung, Doug Szajda, Mark Andersland, Leo Tam, C.P. Watkins, Brian L. Mark, Miguel A. Labrador, Gene Chase, Harry W.Tyrer, Robert Siegfried, Harlan B. Russell, John R. Black, Robert Y. Ling, Julia Johnson, Karen Collins, Clark Verbrugge, Monjy Rabemanantsoa, Kerry D. LaViolette, William Honig, Kevin Mills, Murat Demirer, J Rufinus, Manton Matthews, Errin W. Fulp, Wayne Daniel, Luiz DaSilva, Don Yates, Raouf Boules, Nick McKeown, Neil T. Spring, Kris Verma, Szuecs Laszlo, Ted Herman, Mark Sternhagen, Zongming Fei, Dulal C. Kar, Mingyan Liu, Ken Surendran, Rakesh Arya, Mario J. Gonzalez, Annie Stanton, Tim Batten和 Paul Francis。

其次，在普林斯顿大学的网络系统组的几位成员对本书提供了意见、例子、校订、数据和代码段。我们要特别感谢Andy Bavier, Tammo Spalink, Mike Wawrzoniak, Zuki Gottlieb, George Tzanetakis和Chad Mynhier。正如以前一样，我们感谢国防部高级研究计划署、国家科学基金、Intel公司和思科系统公司在过去几年对我们网络研究课题的支持。

再次，我们衷心的感谢我们的丛书编辑David Clark 以及Morgan Kaufmann出版公司中在本书编写期间帮助过我们的所有人。还要特别感谢我们原来的责任编辑Jennifer Mann，第3版的编辑Rick Adams，我们的制作编辑Karyn Johnson和我们的生产经理Simon Crump。与MKP出版公司全体人员合作的过程很令人愉快。

目　录

第1章 基 础

我必须创造一个体系，不然的话我就会沦为别人体系的附庸；我不要推理也不要比较，我的工作是创造。

——威廉·布莱克

问题：建造一个网络

假设我们要建造一个计算机网，它有潜力发展到全球性的规模，并且能够支持各种各样的应用，如远程会议，视频点播，电子商务，分布式计算和数字化图书馆等。那么要采用什么样的技术作为基础构件，以及使用何种软件体系结构才能把这些构件集成为一个有效的通信服务？本书最主要的目标就是回答这个问题，描述可用的建造材料，以及说明如何自下而上用它们来建造一个网络。

在我们了解如何设计计算机网络之前，首先应在什么是计算机网络这一问题上达成共识。有一个时期，网络（network）一词是指用于将单一功能终端连接到大型计算机所用线路的集合。一些人认为，这个词是指语音电话网络。而另一些人认为，唯一重要的网络是用于传播视频信号的电缆网络。这些网络的主要共同点是专门处理某种特定类型的数据（按键、音频或视频），并且通常连接到特殊用途的设备（终端、手持接收器和电视机）。

计算机网络与其他类型网络有什么区别？也许计算机网络的最主要特征是其通用性。计算机网络主要由通用可编程硬件来构建，并且不会为如打电话或传送电视信号那样的特定应用做任何优化。相反，计算机网络能够运载多种不同类型的数据，并且支持广泛的不断增长的应用。本章考察计算机网络的一些典型应用，然后讨论网络设计者为支持这些应用必须了解的东西。

2

一旦我们弄清楚这些需求，接下来该怎么做呢？幸运的是，我们并不是在建造第一个网络。其他一些人，最著名的是因特网的研究人员群体，已经先于我们完成了这项任务。我们可以从因特网中得到的丰富经验来指导我们的设计。这些经验包含在网络体系结构（network architecture）中，网络体系结构指明可用的软硬件构件，并且说明如何将它们组织起来构成一个完整的网络系统。

为了让我们开始了解如何建造一个网络，本章将完成四项任务。首先，揭示不同的应用和不同的群体（如网络用户和网络操作者）对网络的需求。第二，引入网络体系结构的概念，它是本书其余部分的基础。第三，介绍计算机网络实现的几个关键因素。最后，介绍用来衡量计算机网络性能的关键度量标准。

3

1.1 应用

多数人是通过因特网的各种应用（如万维网、电子邮件、音频和视频流、聊天室和共享音乐文件）来了解因特网的。举例来说，万维网提供了一个直观而且简单的界面。用户浏览有很多文本和图形对象的网页时，点击他们想进一步了解的对象，就会弹出相应的新网页。大多数

人也明白在这个应用的背后，是网页上的每个可选对象都绑定着一个指向下一个页面的标识符。这个标识符被称为统一资源定位器（uniform resource locator, URL），它唯一标识你通过浏览器所能浏览到的每一个网页。举个例子：

```
http://www.mkp.com/pd3e
```

这是指向Morgan Kaufmann 网站上描述本书页面的URL：字符串http表明要下载页面应使用超文本传输协议（HTTP），www.mkp.com是提供网页的计算机的名字，而pd3e唯一标识出版商网站中的该网页。

但是，大多数用户不清楚的是，点击这样一个URL后，在因特网上可能需要交换多达17条消息才能得到网页，并且假定网页本身要小到可以存放在一条消息中。这些消息中有6条用来把服务器名（www.mkp.com）翻译成它所对应的因特网地址（213.38.165.180），3条消息用来建立从你的浏览器到服务器之间的传输控制协议（TCP）的连接，4条消息用来让浏览器发送HTTP的“get”请求，并让服务器回送被请求的页面（以及双方对收到消息的确认）。还有4条消息用来关闭TCP连接。当然，这不算因特网节点一天交换的数以百万计的消息，这些消息只是让节点相互知道自己的存在，并准备提供网页服务，把机器名翻译成网络地址，并将各种消息向它们的最终目的地转发。

尽管现在还不像网上冲浪那样普及，但是音频和视频的流式播放作为因特网的一项应用正在兴起。尽管你可以把整个视频文件从远程主机下载到本地播放，就像下载和显示一个网页的过程一样，但你必须等到视频文件的传送进行到最后一秒才能开始观看。以流的方式播放视频意味着发送方和接收方分别充当流的信源（source）和信宿（sink）。也就是说，信源创建一个视频流（也许是使用视频采集卡），以消息的形式在因特网上发送，而信宿在收到消息后将其显示出来。

4 更准确地说，视频并不是一种应用而是一种数据类型。视频点播是对视频应用一个例子，它先从硬盘上读取一个已有的电影，然后把它传送到网络上去。另一种应用是视频会议。因为它有很严格的时间约束，所以更容易引起人们的注意。就像使用电话一样，参与者之间的交互必须是及时的。当一端的用户做一个动作，这个动作必须尽快被显示在另一端。太长的延迟会造成系统无法使用。相反地，如果从用户打开视频到第一幅图像被显示出来只用了几秒钟，那么这个服务就仍可被认为是令人满意的。此外，交互视频意味着有双向流动的视频数据，而视频点播应用大多只向一个方向发送视频数据。

UNIX应用程序vic是一个流行的视频会议工具。图1-1显示一次vic会话的控制面板。实

5 际上，vic是Lawrence Berkeley 实验室和加州大学Berkeley分校共同开发的一套视频会议工具之一。其他例子包括白板应用程序（wb）（允许用户互相发送草图和幻灯片），一种可视化音频工具vat，和用来创建和发布视频会议的会话目录（sdr）。所有这些在UNIX上运行的工具（它们的名字采用英文小写）都可以在因特网上免费获得。同时还可以获得用于其他操作系统的类似工具。

尽管这只是两个例子，但从网上下载网页和参加视频会议已经能够展示建立在因特网上的应用的多样性，并暗示因特网设计的复杂性。本书余下的部分将从头开始，一次讲述一个问题，解释如何建立一个应用如此之广的网络。第9章作为全书的结束部分将重新讨论这两种特定的应用以及其他几种在当今因特网上流行的应用。

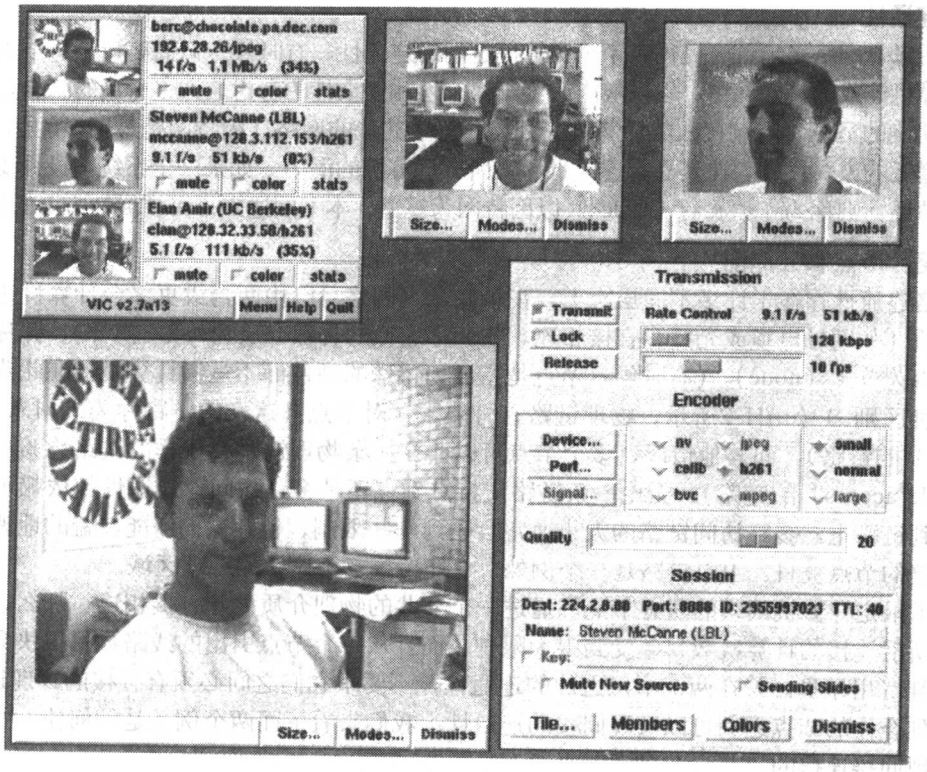


图1-1 vic视频应用

1.2 需求

我们刚刚为自己制定了一个宏伟的目标：即了解如何从最底层开始建造一个计算机网络。要实现这个目标，我们将从基本原理开始，然后提出各种问题，而这些问题在建造实际的网络时会经常遇到。在每一步，我们会用现有的协议去说明各种可行的设计方案，但是我们并不把这些人为的协议当作不变的真理。相反，我们还要不断地提出（并回答）网络为何要如此设计的问题。在力求让人们理解现今网络的工作方式的同时，重要的是认识基本概念，因为随着技术的发展和新应用的不断出现，网络的变化日新月异。根据我们的经验，一旦你理解了基本思想，对于你遇到的任何新协议，消化和吸收起来都将变得相对容易。

首先，要确定影响网络设计的所有约束和需求。但是在开始前，我们要明确一点：你对网络的期望取决于你考虑问题的角度：

- 应用程序员可能会从应用所需的服务来考虑，例如，要保证应用程序发出的每条信息能在一定的时间内准确无误地传递；
- 网络设计者会从设计一个高效率的网络来考虑，例如，有效地利用网络资源和公平地分配给不同的用户。
- 网络提供者则会从系统是否易于执行和管理的特性来考虑，例如，故障易于被隔离，且易于考虑使用情况。

本节力求将这些不同观点提升到一个高的层次，重点介绍驱动网络设计的主要方面，从而明确贯穿全书的各种挑战。

1.2.1 连通性

很明显，网络必须首先提供若干个计算机之间的连通性。有时候，只需要建立一个由选定的几台计算机连成的有限网络。但事实上，鉴于保密性和安全性的考虑，许多专用的（企业的）网络都有明确的目标：限制接入的计算机。相反，一些其他的网络（因特网是最明显的例子），要用一种增长的方式设计，使其具有连入世界上所有计算机的潜力。如果一个系统其设计支持任意规模的扩展，则称为可扩展的（scale）。以因特网为模型，本书阐述来自可扩展性方面的挑战。

链路、节点和云形图

网络连通性存在于许多不同层次上。在最底层，网络可以由两台或更多台计算机通过某种物理介质（如同轴电缆或光纤）直接相连。我们称这样的物理介质为链路（link），并称被连接的计算机为节点（node）。（有时候，节点是指更为特殊的硬件而不是指计算机，在此我们暂时忽略这种区别。）如图1-2所示，物理链路有时限于一对节点（这样的链路称为点到点（point-to-point）的链路），而其他情况，多个节点可以共享一条物理链路（这样的链路称为多路访问（multiple access）的链路）。无论一条链路支持点到点还是多路访问，连通性都依赖于节点如何连接到链路上。多路访问链路的大小通常会受到一些限制，包括它们所能覆盖的地理距离和所能连接的节点数目。卫星链路是一个例外，它能覆盖一个广阔的地理区域。

如果限定计算机网络的所有节点都通过一个公共的物理介质彼此直接相连，那么，网络不仅在它所能连接的计算机数目上受到很大限制，而且从每个节点引出的线路数目很快会变得难以控制且费用昂贵。好在两个节点之间的连通性并不要求它们之间必须有直接的物理连接——在一系列合作的节点之间可以实现间接的连通性。我们来看下面两个例子是如何使一些计算机之间实现间接连接的。

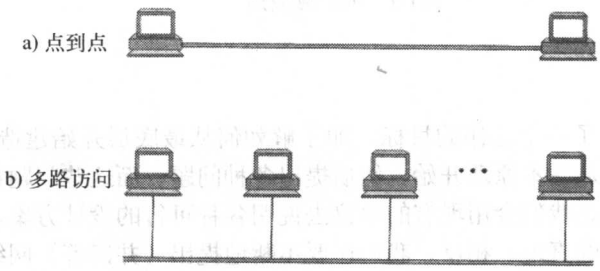


图1-2 直接链路

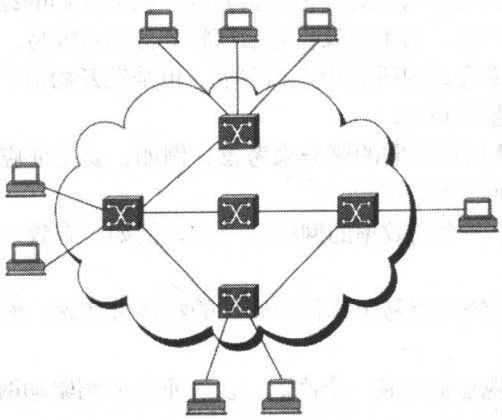


图1-3 交换网

图1-3显示一组节点，每个节点都连到一条或多条点到点链路上。那些连着至少两条链路的节点运行软件，将从一条链路收到的数据转发到另一条链路上。如果按系统化的方法进行组织，这些转发节点形成一个*交换网*（switched network）。有很多种交换网，最普通的两种是*电路交换*（circuit switched）和*分组交换*（packet switched）。前者主要用于电话系统，而后者多用于计算机网络，本书将侧重后者。分组交换网最主要的特点是网络中的节点彼此间发送离散的数据块。可以将这些数据块看作对应于某种应用的数据，如一个文件，一个电子邮件，或一幅图像。我们称每个数据块为一个*分组*（packet）或一条*消息*（message），现在我们可互换地使用这两个术语；但它们并非在任何时候都是一样的，我们将在1.2.2节讨论它们之间的差异。

分组交换网一般使用一种叫做*存储转发*（store-and-forward）的策略。正像其名字一样，在存储转发网络中的每个节点先通过某条链路接收一个完整的分组，将这个分组存于它的内存，然后将完整的分组转发给下一个节点。与其不同的是，电路交换网首先通过一系列链路建立一条专用电路，然后允许源节点通过这条链路发送比特流到达目标节点。在计算机网络中使用分组交换而不使用电路交换的主要原因是效率，这个问题将在下一小节讨论。 8

图1-3中的云形图将实现网络的内部节点（它们通常称为*交换机*（switch），其唯一的功能是存储和转发分组）和云形图之外使用网络的节点（它们通常称为*主机*（host），其任务是支持用户并运行应用程序）区分开来。还要指出，图1-3中的云图是计算机网络中最重要的图标之一。通常，我们用云图表示任何类型的网络。无论网络是一条点到点的链路、一条多路访问的链路或是一个交换网，都可以用云图来表示。因此，无论你在哪个图上看到云图，都可以把它看作本书讨论的任意一种网络技术的表示。

计算机间接连通的第二种方法如图1-4所示。在这种情况下，一些独立的网络（云图）互连形成一个*互连网*（internetwork，或简称internet）。我们按照因特网的习惯，将通称的互连网写成以小写i开头的internet work或internet，而将当前使用TCP/IP的因特网写成大写I开头的 9
Internet。连接两个或多个网络的节点通常称为*路由器*（router）或*网关*（gateway），它与交换机所起的作用大致相同，即把消息从一个网络转发到另一个网络。注意，一个互连网本身可被看作是另一种类型的网络，也就是说，一个较大的互连网可由一些较小的互连网互连而成。这样，我们可以通过将云图互连成更大的云图来递归地建造任意大的网络。

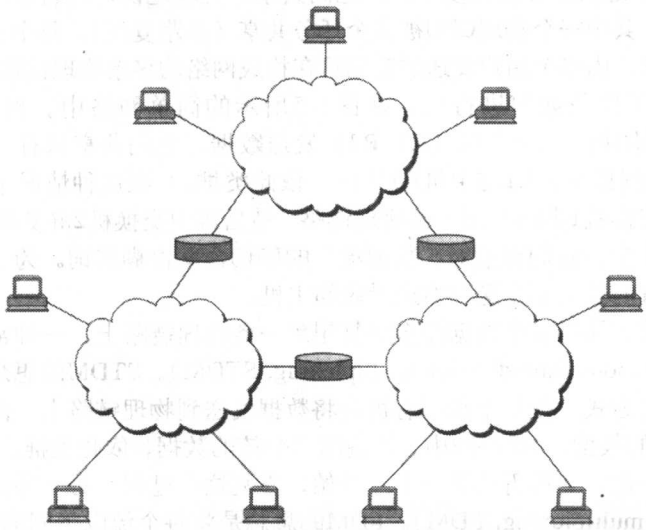


图1-4 网络的互连