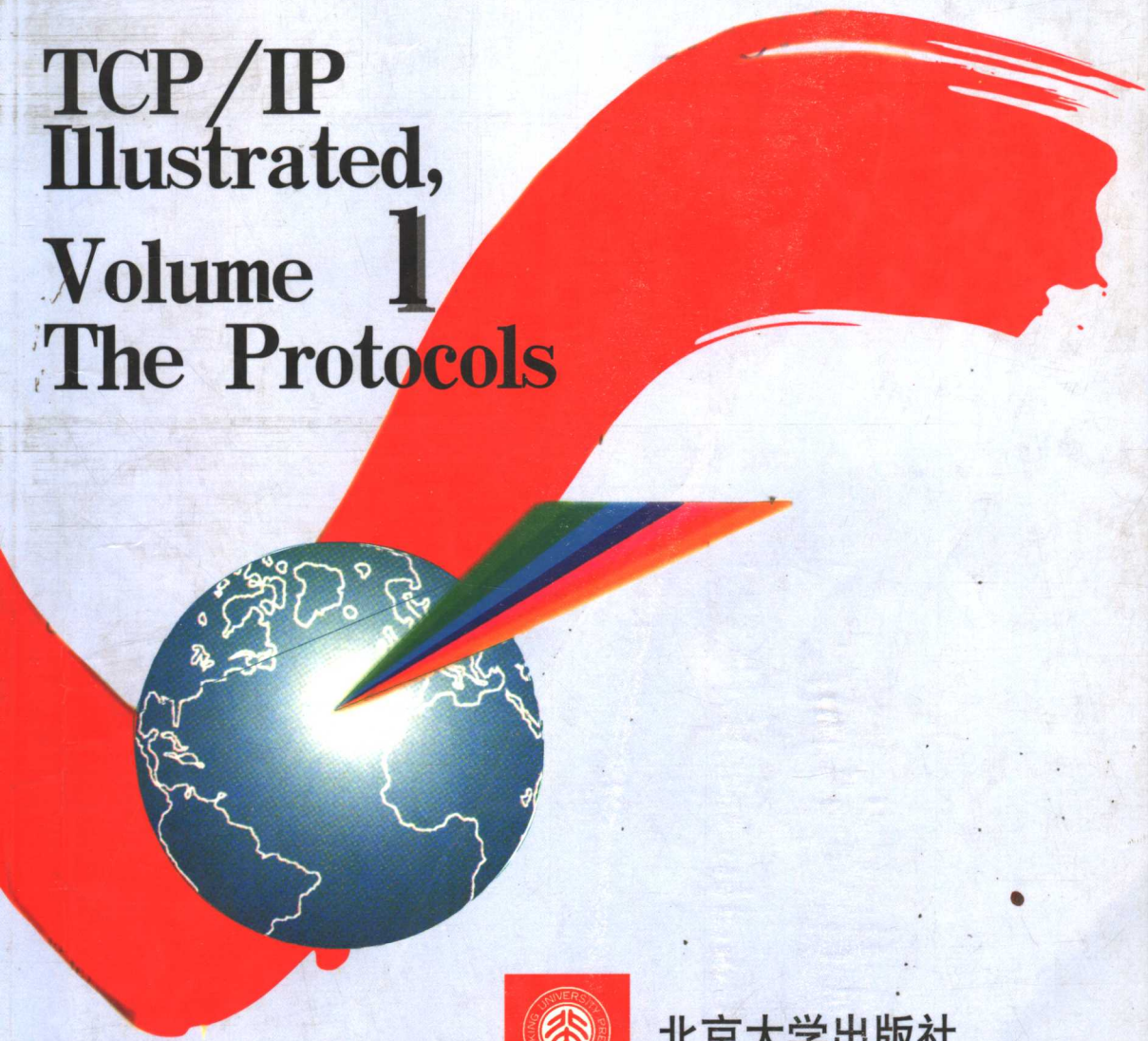
# TCP/IP 详解（第一卷）协议

〔美〕W.Richard Stevens 著

# TCP/IP Illustrated, Volume 1 The Protocols



ADDISON-WESLEY PROFESSIONAL COMPUTING SERIES

北京大学出版社

# TCP/IP 详解

## 第一卷 协议

〔美〕 W. Richard Stevens 著
任守奎 段振海 王 钢 侯克合
王 雷 郭 倩 方兴军 译
任守奎 审校

北京大学出版社
北 京

**著作权合同登记号　图字：01-97-1093**



**图书在版编目(CIP)数据**

TCP/IP 详解　第 1 卷：协议/(美)斯蒂文思(Steve ns. W. R.)著；任守奎等译. —北京：北京大学出版社，1999.1

书名原文：TCP/IP Illustrated

ISBN 7-301-04233-7

Ⅰ.T…　Ⅱ.①斯…　②任…　Ⅲ.计算机网络-通信协议，TCP/IP　Ⅳ.TP393

中国版本图书馆 CIP 数据核字(1999)第 26435 号

**书　　名：TCP/IP 详解(第一卷　协议)**
著作责任者：〔美〕W. Richard Stevens 著，任守奎等译
责 任 编 辑：杨锡林
标 准 书 号：ISBN 7-301-04233-7/TP·469
出　版　者：北京大学出版社
地　　址：北京市海淀区中关村北京大学校内　100871
网　　址：http://cbs.pku.edu.cn/cbs.htm
电　　话：出版部 62752015　发行部 62754140　编辑部 62752032
电 子 信 箱：zpup@pup.pku.edu.cn
排　版　者：兴盛达激光照排中心
印　刷　者：北京银祥福利印刷厂
发　行　者：北京大学出版社
经　销　者：新华书店
787 毫米×1092 毫米　16 开本　29.125 印张　730 千字
1999 年 7 月第一版　1999 年 7 月第一次印刷
定　　价：52.00 元

# 前　言

## 概述

本书讲述了 TCP/IP 协议族，但所选择的角度与其它有关 TCP/IP 的书有所不同。本书并不仅仅描述协议本身以及各自的功能，我们还将使用一个流行的测试工具来观察实际交互中的协议内容。看一看在各种实际环境中协议是如何操作的将有助于我们更好地理解它们如何工作以及为什么那样设计。本书也涉及了协议的实现，但并不要求读者去阅读数千行的源代码。

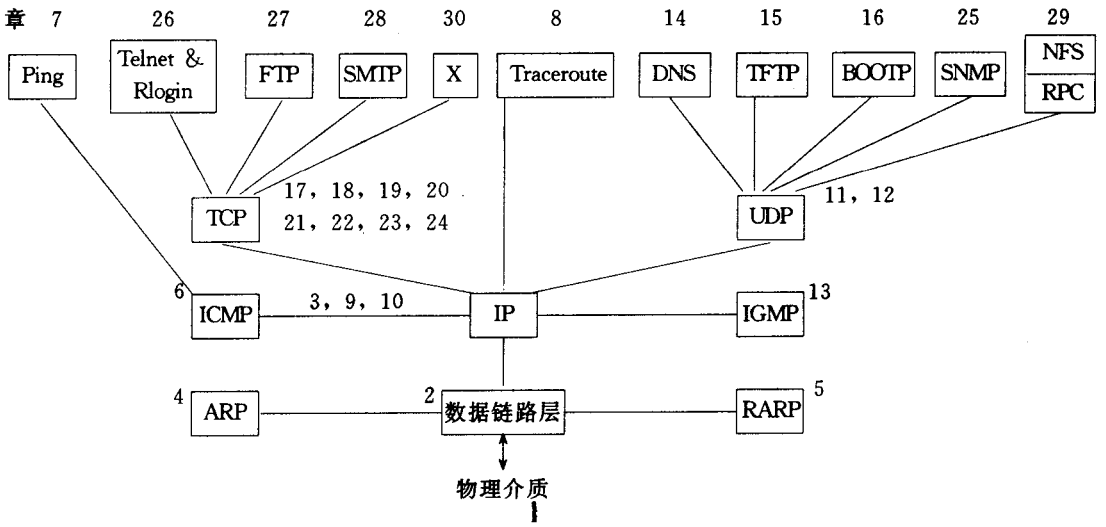
在 60 年代到 80 年代，网络协议正处于被开发的阶段，这时，为了观察“通过缆线”的数据分组，需要专用的昂贵的硬件设备。为了能够理解硬件设备显示的分组内容，必须对协议相当熟悉。而这些硬件分析器的功能也已由硬件设计者限定，不能再由具体用户进行扩展。

现在，这种局面已发生了巨大变化。人们可以用随处可见的工作站来监视一个局域网[Mogul 1990]。将一台工作站连入你的网络，在其上运行一些可公开得到的软件(参见附录 A)，就可以观察缆线上的通信情况。虽然有很多人都认为这类工具是用于检测网络故障的，其实把它们作为理解网络协议如何工作的工具也是很不错的，而这正是本书的目标所在。

本书对于任何希望理解 TCP/IP 协议如何操作的人都是有益的，像网络应用编程人员、负责维护和使用 TCP/IP 协议的计算机系统及网络的系统管理员、经常使用 TCP/IP 应用的一般用户等。

## 本书的内容安排

下图给出了本书中所涉及的协议和应用。方框旁的数字表明该协议或应用是在哪一章被讨论的。



上图并未包含本书的全部内容。例如，对于使用 TCP 的 DNS 和 RPC 情况，这里没有给出相应的图示。

本书由底层向高层来讲解 TCP/IP 协议族。在第一章对 TCP/IP 做了概括介绍之后，从第二章开始讲解链路层并且沿着协议栈逐层向上讲解。这样做可为不熟悉 TCP/IP 协议或缺少一般网络概念的读者提供必要的基础知识。

本书在讲解时注意到了功能方面的关系，而不是刻板地按照由底到高的层次进行。例如，在第三章中讨论了 IP 层以及 IP 报头。但是 IP 报头中的许多域段更适合于在使用这些域段或受这些域段影响的应用程序中讨论。例如，由于 UDP 数据报经常受到分段的影响，所以在讲解 UDP(第十一章)协议时再来讨论分段。我们将在第八章中讨论 Traceroute 应用时来讲解生存期(time-to-live)这个域段，该域段是 Traceroute 操作的基础。同样地，许多 ICMP 的特性也放到了后面的章节中讨论。

并非所有好东西都拖到最后才讲，当读者具备了理解 TCP/IP 应用的基础时，我们就对之加以讨论。在讲述了 IP 和 ICMP 之后，我们将讨论 Ping 和 Traceroute 应用。在讨论了 UDP 协议之后，我们将讲解基于 UDP 的应用(多点广播，DNS，TFTP 和 BOOTP)。TCP 应用以及网络管理将要放到最后去讨论，本书将侧重于这些应用如何使用 TCP/IP 协议而不是运行这些应用的细节。

### 读者

本书由浅入深地进行讲解，不需要读者具有网络或 TCP/IP 的专门知识。本书也提供了大量的参考文献以供对某些专题有兴趣的读者阅读。

对本书可以有多种阅读方法，它可以被当作自学参考书，从头到尾对整个 TCP/IP 协议族进行学习；而具有一定 TCP/IP 知识的读者可以跳过前面几章，直接从第七章开始阅读，并且主要集中于自己感兴趣的章节。在每章的后面都附有练习题，附录 D 给出了大部分练习题的答案，这对于自学者尤为有用。

当被用作一个学期或两个学期的计算机网络课程的教科书时，重点应当放在 IP(第三章和第九章)，UDP(第十一章)，TCP(第十七至二十四章)以及一些有关应用程序的章节。

为了使得某些章节适于单独学习，本书中给出了许多向前和向后的参照指示，以及一个贯穿全书的索引。

如果你有机会上网，我们希望你能得到本书中所用的应用软件(附录 F)，并且自己动手试一试，这样学习可以获得更多的知识。

### 本书中所用到的测试系统

本书中的所有例子都运行在实际的网络环境中，而且是将运行结果保存在文件中以便于本书使用。图 1.11 给出了本书所用到的不同主机、路由器以及网络的示例图。图中这个网络拓扑结构较为简单，易于理解。其中有四个系统用作路由器，由此可以观察路由器产生的错误报文。

大多数系统的名称反映了所使用的软件类型，如 bsdi，svr4，sun，solaris，aix，slip 等等。这样通过系统名就可以知道正在打交道的软件类型。

本书中用到了许多不同的操作系统以及 TCP/IP 实现：

● BSD/386 版本 1.0，来自 Berkeley 软件设计公司。安装在主机 bsdi 以及 slip 上。这个系统是由 BSD 网络软件 2.0 版派生来的。（在图 1.10 中我们给出了各种 BSD 版本的族系关系。）

● Unix System V/386 Release 4.0 Version 2.0，来自 U.H. 公司，安装在主机 svr4 上。这是一种较常见的 SVR4，它包含了大多数 SVR4 版本都使用的来自 Lachman 协会的标准 TCP/IP 实现。

● Sun OS 4.1.3，来自 Sun 微系统公司，安装在主机 sun 上。SunOS 4.1.x 可能是使用最广泛的 TCP/IP 实现，其 TCP/IP 代码来自 4.2 BSD 和 4.3 BSD。

● Solaris 2.2，来自 Sun 微系统公司，安装在主机 solaris 上。solaris 2.x 中的 TCP/IP 实现不同于早期的 SunOS 4.1.x 系统以及 SVR4。（这个操作系统实际是 SunOS 5.2，但通常都称其为 Solaris 2.2。）

● AIX 3.2.2 来自 IBM，安装在主机 aix 上。其 TCP/IP 实现是基于 4.3 BSD Reno 版的。

● 4.4 BSD 来自美国加州大学伯克利分校的计算机系统研究小组，安装在主机 vangogh.cs.berkeley.edu 上。这个系统具有最新的来自 Berkeley 的 TCP/IP 实现。（这个系统并没有在示例图中给出，可以通过 Internet 进行访问。）

虽然上述系统都是 Unix 系统，但 TCP/IP 实现是和操作系统无关的，几乎所有流行的非 Unix 系统也都支持 TCP/IP 协议。本书的大部分内容对于非 Unix 实现也是适用的，尽管可能有些程序（例如 Traceroute）不是被所有系统支持。

**排版约定**

当显示交互式输入和输出时，我们用黑体表示键盘输入，而计算机输出用正常字体，注释行用斜体加在旁边，如下所示：

```
bsdi % telnet svr4 discard      connect to the discard server
Trying 140.252.13.34...         this line and next output by Telnet client
Connected to svr4.
```

另外，我们将系统名用作 shell 提示符的一部分（本例中为 bsdi）来显示该命令是在哪个主机上运行。

*在本书中，我们还将使用缩进的插入注释（像本行）来描述历史发展或实现细节。*

有时我们将引用 Unix 手册中有关命令的描述，例如 ifconfig(8)。命令名后跟括号中的数字是引用 Unix 命令的常用方法。括号中的数字是 Unix 手册中的章节数。不幸的是，并不是所有的 Unix 系统都使用相同的章节号。我们将使用 BSD 风格的章节号（与 SunOS 4.1.3 等从 BSD 派生来的系统具有相同的章节号），可能与你的系统不同。

**致谢**

虽然只有作者本人的名字出现在封面上，但一本好书的出版是许多人通力合作的结果。最重要的是作者的家人，他们陪伴着我度过了漫长的著书时间。再次感谢你们：Sally，Bill，Ellen 以及 David。

资深编辑 Brian Kernighan 无疑是非常出色的。他第一个阅读了本书的许多草稿，并用红笔对其中的内容进行了标定，他对细节的留心，他对文法的讲究使作者得到很大的好处。

技术审查人员使作者的疵误得到指正。他们的评论、建议以及（更重要的）批评使本书增色

不少。Steve Bellovin，Jon Crowcroft，Pete Haverlock，以及 Doug Schmidt 阅读了整个初稿，Dave Borman，Tony DeSimone，Bob Gilligan，Jeff Gitlin，John Gulbenkian，Tom Herbert，Mukesh Kacker，Barry Margolin，Paul Mockapetris，Burr Nelson，Steve Rago，James Risner，Chris Walquist，Phil Winterbottom，以及 Gary Wright 阅读了部分章节，Dave Borman 阅读了所有 TCP 章节，Bob Gilligan 实际上和作者一起提供了附录 E。感谢他们。

Joe Godsil，Jim Hogue，Mike Karels，Paul Lucchina，Craig Partridge，Thomas Skibo，以及 Jerry Toporek 给了我不少帮助，他们不厌其烦地回答了我的许多 e-mail 问题。

我经常被问到许多有关 TCP/IP 的问题，而其中一些我不能立即给出答案。这使我认识到进行一些测试可能是回答这些问题的较好方法。感谢 Pete Haverlock 询问有关测试问题，感谢 Van Jacobson 提供了许多本书中用到的公共测试程序。

写作一本有关计算机网络的书需要一个真实的网络环境。感谢国家光学天文台(NOAO)提供的网络环境，感谢 Sidney Wolff，Richard Wolff，以及 Steve Grandi 允许我访问他们的主机，特别要感谢 Steve Grandi 回答了我许多问题并在多台主机上为我提供账号。另外还要感谢美国加州大学伯克利分校的 Keith Bostic 以及 Kirk McKusick，使我能够访问最新的 4.4 BSD 系统。

最后，是出版社的努力使该书得以和读者见面。特别感谢 John Wait，与他以及 Addison-Wesley 出版社的其他专家一起工作是一件令人愉快的事情。

本书的胶片影印拷贝是由作者使用 James Clark 的 Groff 包制成的。我期待着读者的评论、建议以及对其中错误的纠正。

*Tucson, Arizona*
*October 1993*

W. Richard Stevens
rstevens @ noao.edu
http://www.noao.edu/~rstevens

# 目　录

# 第一章 概 述

## 1.1 简 介

TCP/IP 协议族可以允许众多不同厂家的、运行不同操作系统的各类计算机进行相互通信。它的使用规模大大超过了最初的估计，这不能不说是一件令人惊奇的事情。这项工作始于60年代末，最初是由政府资助来进行分组交换网络的研究，而到了90年代，它已发展成为使用最广泛的计算机互连形式。TCP/IP 协议族是一个真正的开放系统，它的协议定义以及大量实现都可以免费或花少量钱得到。称之为世界范围的因特网或简称因特网的网络已经发展成为真正意义上跨越全球的拥有百万台以上计算机的广域网，而其基础正是 TCP/IP 协议族。

本章将对 TCP/IP 协议族进行一次概括性的介绍，为后面的章节提供基本的背景知识。如果想对 TCP/IP 的早期发展作进一步了解，请参阅[Lynch 1993]。

## 1.2 分 层 模 型

网络协议一般都分为若干层次，每一层负责通信的不同方面。一个协议族，像 TCP/IP 等，是各层不同协议的集合体。如图 1.1 所示，TCP/IP 一般被认为是一个四层系统。

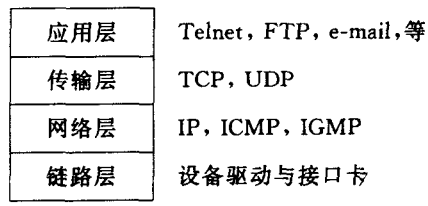



**图 1.1 TCP/IP 协议族的四层模型**

每一层实现不同的功能。

1. 链路层，有时也称为数据链路层或网络接口层，一般包括操作系统中的设备驱动程序以及计算机中相应的网络接口卡。两者接合起来一同处理与电缆线（或使用的其它类型的介质）接口的物理硬件细节。

2. 网络层（有时也称为 internet 层）处理网络中分组的移动。例如，分组的路由选择就是在该层实现的。IP（网际网协议）、ICMP（网际控制报文协议）和 IGMP（网际组管理协议）构成了 TCP/IP 协议族中的网络层。

3. 传输层为其上的应用层提供了两主机之间的数据流服务。在 TCP/IP 协议族中有两类非常不同的传输层协议：TCP（传输控制协议）和 UDP（用户数据报协议）。

TCP 协议提供了两主机之间的可靠数据流。它的功能包括：将应用层传递给它的数据分割为适于网络层的数据块，接收分组确认，为发送的分组设置超时间隔以保证对方收到该分

组，等等。因为传输层提供了可靠的数据流，应用层就可以不必再考虑这些细节。

而 UDP 则只为应用层提供了相当简单的服务。它只负责向对方发送称为数据报的数据分组，而不保证这些数据报可以到达对方，必须由应用层自己来实现它所需要的可靠机制。

每种不同的传输层协议都有其应用，这些我们将在后面的章节中看到。

4. 应用层处理特定的应用细节。有许多 TCP/IP 应用服务在几乎所有的实现中都提供：

- Telnet：远程登录服务；
- FTP：文件传输协议；
- SMTP：简单邮件传输协议，电子函件服务；
- SNMP：简单网络管理协议；

以及其它一些服务，这些我们将在后面的章节中详细介绍。

假定我们在某局域网（像某以太网）上有两台主机，图 1.2 给出了运行 FTP 服务时所涉及的协议。

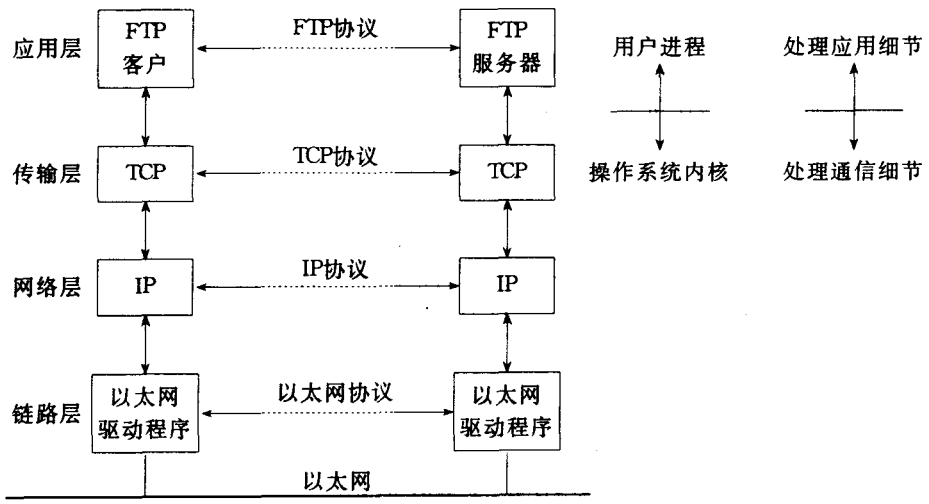



**图 1.2　LAN 上运行 FTP 的两主机**

在图 1.2 中，我们将两个应用框一个标记为 FTP 客户，另一个标记为 FTP 服务器。大多数网络应用也都是这样来实现的：一方为客户，另一方为服务器。服务器为客户提供某种形式的服务，在本例中即为对服务器主机文件的存取。在远程登录 Telnet 服务中，服务器为客户提供登录到服务器主机的能力。

每一层都有一个或多个协议用于与同层的对等方通信。例如，一个协议允许两个 TCP 层进行通信，而另一个协议用于两个 IP 层进行通信。

在图 1.2 的右侧，我们指出应用层是用户进程，而下三层是在内核（操作系统）中实现的。虽然并不是必须这样做，但这种实现具有典型性，而且在 Unix 环境中它就是这样实现的。

在图 1.2 的最上层与下三层之间还有其它重要的不同之处。应用层只关心应用服务的细节而不考虑数据在网上是如何移动的。而下三层只处理通信细节，对具体应用一无所知。

在图 1.2 中，我们共显示出了 4 种协议，每层一种：FTP 是应用层协议，TCP 是传输层协议，IP 是网络层协议，而以太网协议是链路层协议。TCP/IP 协议族是众多协议的集合体。虽然