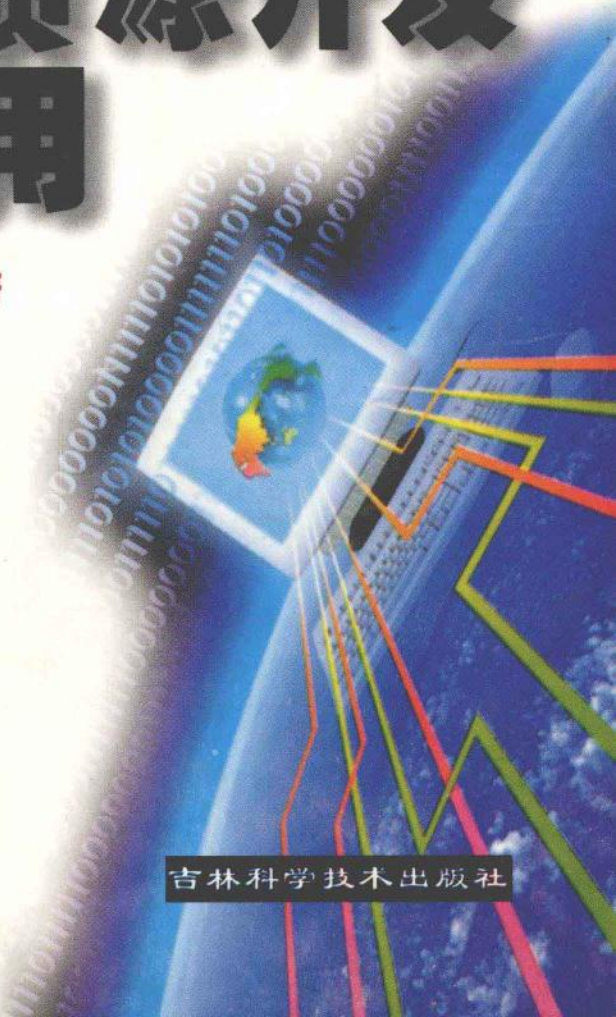
internet

信息资源开发与利用

毕 强 刘春茂 著



吉林科学技术出版社

Internet 信息资源开发与利用

毕　强　刘春茂　著

吉林科学技术出版社

[吉]新登字 03 号

Internet 信息资源开发与利用　　毕　强　刘春茂　著

责任编辑　赵玉秋　　封面设计　李彬彬

出版发行　吉林科学技术出版社　开本 850×1168 毫米　1/32　11.2 印张
292 千字
2001 年 1 月第 1 版　2001 年 1 月第 1 次印刷
印数　1—1300　　定价　20.00 元
印刷　原吉林工业大学印刷厂　　ISBN 7-5384-1886-5/F.188-4

前　言

Internet 网络是信息时代的宠儿。Internet 网络的诱聚吸引力来自它的“二高一低”——极高的扩容速度和极高的参与随意性、极低的表达束缚力。网络经济的先导性和优越性，大大拓展了人类利用信息资源的广度和深度，改变了传统经济学“资源有限，需求无限”的矛盾。网络信息资源不仅是新时代社会生产力的集中体现，而且也预示着社会生产关系面临重大的变革。如果说资本是工业社会的主角，那么网络信息资源将是信息社会的中坚。网络信息资源的可持续开发，知识和信息本位制的有效建立，使信息管理建立在知识和网络信息资源的基础上成为可能，客观上要求信息管理把网络信息资源开发与利用作为重要的研究对象。这正是本书的目的。

针对网络环境下信息资源开发与利用中关键性的信息分布、信息组织、信息发布、信息检索和信息服务等问题和我国在网络信息资源开发与利用中的现实需要，本书力求从多侧面、多视角对这一主题领域进行比较全面和深入的分析和探讨，应该承认，对网络信息资源开发与利用研究的难度较大。由于这是一个全新的研究领域，技术发展和更新的速度非常快，涉及的范围极广泛，所以，围绕这一领域目前可供借鉴的权威性、有一定深度、完整的研究成果尚不多见，本书试图在这方面作一尝试，并侧重分析论证 Internet 信息资源开发与利用的策略，注重操作性和谋略性。

本书共分七章，主要内容包括：(1)Internet 运行基础与 Internet

资源；(2) Internet 信息资源分布与搜集；(3) Internet 信息资源组织；(4) Internet 信息资源的发布；(5) Internet 信息资源检索；(6) 电子信息服务；(7) 面向内容的 Internet 信息资源开发与利用的新技术。在编写过程中，力求简练、结构体系新颖，注重理论与方法的结合，立足用户和读者，尽量避免陈旧过时的内容，突出该领域的最新研究成果和进展。

本书的编写分工如下：刘春茂负责第五章第三节、第五节和第六章；其余各章节由毕强负责。全书最后由毕强修改定稿。

在本书的资料收集和编写过程中得到了天津师范大学信息产业学系系主任刘春茂同志和东北师范大学图书馆网络中心我的学生们的友情支持。原吉林工业大学经济管理学院院长靖继鹏教授给了我热情的鼓励和支持。需要特别指出的是，在编写过程中，作者参阅了大量的国内外资料和案例，在参考文献中并未一一列出，谨在此向他们表示由衷的敬意和感谢！

由于编著者水平有限，加之时间仓促，疏漏及缺点、错误在所难免，敬请批评指正。

毕　强

2000 年 12 月于吉林大学

目　录

第一章　Internet 运行基础与 Internet 资源

当今社会，网络技术已逐步成为现代信息技术的主流。网络的概念也随其技术与应用的迅猛发展而渐入人心。与此同时，计算机的概念也由原始的分立式走向今天的网络式。Internet 就是最好的体现。

1.1　Internet 的运行基础

1.1.1　Internet 的组成

Internet 亦称国际互联网。按字面之意是网络的网络，即网络互联之意。网络互联是网络与网络用相同的通信协议和标准或在网络间增设一接口装置进行标准转换，再用一定的媒体连接（如图 1-1）。

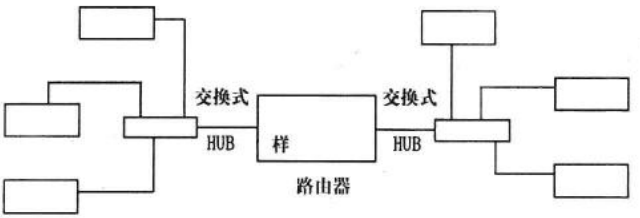


图 1-1　两个计算机网络互联图

从物理结构上看，网络互联又可定义为在网络通信协议控制下，由若干计算机、终端设备、数据传输设备和通信控制处理机等组成

的系统集合，该系统的任何一方均可将它的网络再与其它网络互联，这样便形成了这种国际互联的局面。Internet 网络就是世界上大大小小各种计算机网络互联的集合体，即 Internet＝网络联合国。

从另一个角度来看，所有人的参与，使得 Internet 成为宝贵的信息资源网。因此，中国人将其称为“全球信息资源网”，而西方人则称 Internet 为“A Wonderful World”（一个极妙的世界）。

一般认为 Internet 是由五大部分组成：①人；②信息资源；③计算机；④网络互联设备；⑤通信线路。

1.1.2 带宽

带宽指的是一条通信线路传输数据能力的高、低，或者说通信线路的速度。我们可以这样理解带宽，即一条通信线路每秒钟可以传送多少个二进制的位，这很类似于一条公路可以并排走多少辆汽车，或是一条排水沟一秒钟可以通过多少立方米的水。带宽越大，网络的效率就越高。带宽的基本量和表示单位是 bps（bits per second）。例如，质量好的电话网的带宽可以达到 56Kbps 甚至更高；一条 DDN 专线（数字数据网）的带宽可以达到 2Mbps；一条 X.25 的虚电路可以达到 64Kbps；普通的以太网可以提供 10Mbps 的带宽；快速以太网可以达到 100Mbps；ATM 网络可以提供 622Mbps 甚至几个 G 的带宽。

带宽是网络的重要资源，如果带宽大，我们就可以在传输数据时省许多时间；如果带宽足够大，我们就可以利用网络去做一些对实时性要求较高的事情，比如在 Internet 上看现场直播等。

1.1.3 路由器

路由器是一种网间连接设备，它最基本的任务是把数据从一个网络送到另外一个网络，这样经过很多路由器的传递，数据就能够正确地到达它的目的地。路由器是一种包交换设备，由于 Internet 使用 IP 协议，因此，Internet 中的大部分路由器被称为“IP 路由器”。IP 路由器在 Internet 中的作用类似于邮局，当你把一封信投递给邮

局后，邮局的工作人员会根据你的投递地址来决定下一步将信送往哪一个邮局，而下一个邮局也会根据一定的原则将你的信继续向下投递，最终你的信会到达收信人的手中。

从功能上看，路由器具有两大功能：①建立路由表；②转发 IP 数据报。

目前大多数的企事业单位和部门连 Internet 网，通常都是一台路由器与 ISP 连接实现。这台路由器就是沟通外部 Internet 和内部网的桥梁。如果网站系统为宽带网可通过两台路由器，一台路由器接入宽带网，用于外网路由并起着防护的作用；另一台路由器用于内网与外网的隔离，以提高系统的安全性。现在大多数的路由器都是 Cisco 公司的产品或与其功能近似。

1.1.4 通信线路

通信线路是 Internet 的基础设施，各种各样的通信线路把 Internet 的数万个路由器连接起来，把用户的计算机与路由器连接起来，可以说，没有通信线路，就没有 Internet。我们对通信线路的最大期望是：更大的带宽、更高的通信质量保证、更低的误码率、更加广泛的地域覆盖。从目前的技术发展角度看来，光纤、卫星、铜线、无线电等通信线路正被广泛应用在网络互联和 Internet 中。

现在我国已有六条专用通讯线路作为国际出口进入 Internet。这六条专线是：

(1) 中国科学院高能物理所网的出口专线

1994 年 4 月，美国政府正式同意 CICICO 路由器输入中国。同时中科院高能所获准使用美国能源网——ESNET，1994 年 5 月高能所正式进入 Internet，速率为 64Kbps。1994 年 7 月，高能所专线改由日本国际电信局的 64Kbps 卫星频道连接日本国家高能物理实验室 (KEK)，再由 KEK 以 512Kbps 的速率去美国。

(2) 中关村网络中心的出口专线

中关村网络中心是由中科院、清华大学和北京大学三单位合建

的教育研究示范网。该网于 1994 年 7 月经美国 SPRING 卫星进入 Internet，速率为 128Kbps。

(3) 中国教育科研网的出口专线

中国教育科研网，简称 CERNET，速率为 2Mbps。它的网管中心设在清华大学，其子网分布在全国七大区的八大城市中，共有十个主节点：

①华北地区：在北京有三个主节点，分别设在清华大学、北京大学和北京邮电大学。

②东北地区：在沈阳有一个主节点，设在沈阳工业大学。

③华中地区：在武汉有一个主节点，设在华中理工大学。

④华南地区：在广州有一个主节点，设在华南理工大学。

⑤华东地区：在上海和南京各有一个主节点，分别设在上海交通大学和南京东南大学。

⑥西南地区：在成都有一个主节点，设在成都科技大学。

⑦西北地区：在西安有一个主节点，设在西安交通大学。

(4) 北京化工大学网的出口专线

(5) 邮电部网出口专线 (CHINANET)

(6) 电子部网出口专线 (商业网)

1.2 Internet 的运行方式

Internet 运行与管理方式有三种：主干网、区域网和局域网。局域网，它是 Internet 的基础模块。Internet 通过它将科研机构、高等院校、公司企业以及其他组织机构内的计算机进行联接，形成数以万计的局域网。区域网，它是由数量不等的局域网通过线路互联而成的。因此，区域网是许多不同的局域网和组织的联合体。它通常连接在全国和洲际的主干线上。主干网，它是由数量不等的区域网互联而成的。Internet 网中用户将数据送到网上 (Upload file)，其数据的运行方式为局域网－区域网－主干网；用户从 Internet 网中

下载（Download file）数据，其数据的运行方式为主干网－区域网－局域网，Internet的这种运行方式又称为交互式三级结构。

1.2.1 局域网络技术选型

从应用层角度组建计算机网络系统，主要包含：网络的结构化布线系统和网络互联与管理两大部分。

（1）网络结构化布线系统

网络的结构化布线系统应在设计建筑物综合布线系统PDS（Premiss Distribution System）时统筹考虑。理想的PDS应支持话音、数据、图像和视频等多媒体信息的综合应用。目前，国际上和我国均已制定出相应的标准，设计时，必须符合这些标准。PDS目前可划分为六个子系统，其中，垂直主干和水平布线子系统，通常分别采用光纤和CAT5类非屏蔽双绞线，可适应于100Mbps快速以太网和155MbpsATM。

（2）局域网（LAN）拓朴结构

通常，在通信网络中，“拓朴”是指网络上的端系统或站点之间的互联方式。局域网中常见的拓朴类型有总线型、环型和星型等。每种拓朴都有其优缺点。环型拓朴是由计算机组成的一个闭环，具有提供较大吞吐量的潜力，并易于检测和协调网络的运行。但是，如果其中一条链路或一个转发器的故障会导致整个网络失败。总线型拓朴结构灵活、易于扩展，它包含一条共享电缆，可适应大范围的布线。但是，这种“共享介质”的方式，可能出现“数据冲突”造成传输失败。星型拓朴具有在一个建筑物中自然布线的优点，并能够保护网络不受一根电缆损坏的影响，但中心节点的故障可能造成全网瘫痪。常见的局域网实例有：总线拓朴结构的以太网（Ethernet），环状拓朴结构的IBM令牌环网和FDDI（Fiber Distributed Data Interface）网；星型拓朴结构的ATM（Asynchronous Transfer Mode）网。选择时应综合考虑系统的实际应用需求，各种局域网的性能特点和价格等诸因素确定。

FDDI 是 Fiber Distributed Data Interface 的缩写，中文意思是光纤分布式数据接口。FDDI 由两条反向传输的光纤组成。FDDI 技术成熟，具有高速（100Mbps）、可靠、冗余和故障恢复等特点，但其分布较为分散，不适合中心管理的模式，而且价格较为昂贵。

令牌环网传输速率有限（16Mbps），其技术发展缓慢，且设备价格较高。

近年来，以太网技术向高速和交换的方向迅速发展，具有灵活性、伸缩性和优良的性能价格比等优点。快速以太网技术将以太网从传统的 10Mbps 传输速率提高到了 100Mbps，更有千兆以太网（Gigabit）将传输速率提高到了 1000Mbps 以上。交换式以太网是以太网技术中的一种，它大大增加了网络的整体带宽。

ATM 技术融汇了高速、交换技术的优点，其传输速率可达 155Mbps，远程 ATM 可达 655Mbps～2.5Gbps 的传输速率；不仅如此，它的带宽特征，实现了单一线路中集成传输数据、语音、图像和音频等多媒体信息的潜力。特别适用于具有大量用户和多媒体应用的大型网络。但是，采用 ATM 技术需要用光纤来代替铜缆，其组网价格昂贵。

表 1-1　几种典型高速网络性能比较

网络类型	优　　点	缺　　点
交换以太网	协议熟悉、支持广泛，升级容易	带宽仍然有限
快速以太网	协议熟悉、支持广泛、后向兼容	结点增加，发展潜力不大
FDDI	技术成熟，使用比较普遍，适用于干线	价格昂贵，发展潜力不大
千兆以太网	协议熟悉、支持广泛，可实现无缝升级	传输质量不能保证，会出现数据包丢失
ATM	扩展性好，可以综合语音、影像、数据、支持实时传播，发展潜力大	价格昂贵、技术复杂、配置和实现复杂

综上所述，中小型网络系统宜采用快速以太网技术，交换式集线器（HUB）与工作站之间可以 10/100Mbps 传输速率连接；将来可根据需要升级为千兆以太网。此外交换技术的发展为虚拟局域网的实现提供了技术基础。

（3）广域网（WAN）接入 Internet

通常，广域网接入 Internet 的方法有多种，不同方法的费用和效率各不相同。广域网接入 Internet 的方式有微波通信、ISDN、ADSL 非对称数字用户光纤线、DDN 数字数据网或帧中继（FR）以及电话专线等几种方式。

①拨号接入 Internet

拨号接入是指通过普通模拟电话线接入中国公用计算机互联网（CHINANET）。CHINANET 通过高速数字专线与国际计算机互联网互联，采用调制解调器（俗称“猫”）连接，提供 56K 的传输速率。此种方式的特点是：安装方便、费用低廉、维护简便，适合大多数家庭使用。拨号接入码又分为 163 与 169 种，拨 163 可访问全球所有的站点，拨 169 只可访问国内的站点。

②专线接入 Internet

通过特殊的接入设备和线路连接到国际互联网（Internet）上。专线方式的基本特征是用户与 Internet 之间保持着相对永久的连接，用户在访问时不需要有类似于电话拨号的呼叫或是连接过程。

一般来说，需要通过专线接入 Internet 的用户大致有五种原因：

●希望与 ISP 保持永久的通信联系；

●希望在 Internet 上提供一些信息服务；

●希望在单位内部设置与 Internet 连接的电子邮件服务器。如有些单位希望给每个员工都分配一个 E－mail 地址，如果 E－mail 地址的需求量过大，向 ISP 申请就不太合算，而且很多单位不希望由 ISP 来保管自己的电子邮件。这样的单位往往在自己的内部网络设置一个 E－mail 服务器，并且使自己的网络与 Internet 保持永久

的连接；

●希望通过 Internet 实现内部网的互联；

●希望自己到 ISP 之间能够有更多的带宽。

现在比较常用的方式有以下两种：

●ISDN (Integrated Services Network) 即综合业务数字网，俗称“一线通”，是一种集话音、数据于一身的电话交换网络。ISDN 是从普通电话交换技术发展而来，在很多地方类似于电话网。例如用户在通过 ISDN 通信时也需要一个拨号过程（需要拨打的通常是 ISDN 号码而非电话号码），而且 ISDN 本身也是电路交换网络。但 ISDN 与普通电话交换网有很多本质区别，如 ISDN 是数字网络而非模拟网络。

ISDN 用户得到的通信带宽要大于电话网，国内的 ISDN 网采用一种被称为“2B+D”的模式，即用户可以使用 2 个 B 信道和一个 D 信道，每个 B 信道的带宽可以达到 64Kbps，一个 D 信道可以达到 16Kbps，但 D 信道通常不用来传送用户的信息，因此 2B+D 的用户可以得到 128Kbps 的带宽。由于两个 B 信道和一个 D 信道，因此用户可以同时与三个不同的方向通信（数据或话音）。ISDN 接入 Internet 适合于公司（小型网络用户）及部分家庭使用。

●DDN (Digital Data Network) 即数字数据网（通常是指中国电信的 CHINADDN），是一种数字传输网络，DDN 将用户的各种速率的数据信息汇集起来，在网络中的速率更高（上网速度可达 64K 到 2M）、质量更好的数字信道上传输。DDN 由数字信道、DDN 节点、网络管理和用户环路等主要环节组成，DDN 各个节点之间的物理通信线路通常采用光纤、数字微波或是卫星信道。纯粹意义上的 DDN 只向用户提供点到点和点到多点（广播或轮询）的透明数字信道，因此适合于 DDN 的用户利用 DDN 线路资源开设多种业务（银行、证券公司），例如 DDN 上的帧中继业务、X.25 业务、会议电视业务甚至普通电话业务等。DDN 最重要的特点就是它的数字电路，

比电话网提供的模拟专线业务传输质量更高、传输延时更小、传输的可靠性更高。

此外，DDN 用户需要关心的问题有①DDN 专线的带宽是多少；② ISP 需要为用户做什么。对于 DDN 的数字专线，用户和 ISP 是对等的两个点，因此申请这条 DDN 专线是双方的事情而不单单与用户相关。用户需要申请通向 DDN 节点（或相关复用设备）的模拟专线，需要购买调制解调器，ISP 也需要花钱去办理这些事务。

（4）网络安全管理

网络的安全性已成为当今网络管理者关注的主要问题。TCP/IP 协议是因特网的核心，其设计目标是开放性而不在于安全性。信息在因特网上传输，面临各种威胁。我国和世界各地已屡遭计算机病毒和黑客的侵扰。由于计算机网络犯罪智能化的特点，使其更具有危害性。从现有的技术手段看，服务器本身不能提供有效的防护措施。只有从网络结构设计入手，提高整个网络的安全性。路由器具有一定的过滤和隔离作用，但其设计的出发点是连接不同的网络，而不是分隔网络。而且，由于控制路由器可以有多种途径，如果其中一种失控，路由器就会被攻破，使网络处于被攻击范围内。为了防止网络遭受非法入侵和内部信息的非法泄露，必须在本地 Intranet 和 Internet 网之间采用防火墙技术，以及与之配套的加密、身份验证、数字签名和内容检查来保证本地网络的安全。防火墙技术已成为广域网设计中的关键技术之一。防火墙是一个实现安全策略的系统或系统组，强制执行对 Intranet 和 Internet 的访问控制，它能保证只有授权的人可以访问 Intranet，且保护其中的资源和有价值的数据不会流出 Intranet。简单的说，防火墙就是介于两个网络之间的具有某些存取控制功能的软硬件集合，它的主要目的是控制数据组，只允许合法流通，从而达到保障网络安全的目的，防火墙技术假设被保护网络具有明确的边界和服务，并且假设网络信息的威胁主要来自外部网络而不是内部网络，它通过建立一整套规则和系统策略来

检测、限制、更改穿越防火墙的数据流，实现对内部网络的保护。防火墙一般主要包括五部分：安全操作系统，过滤器，网关，域名服务和 Email。常见有：Filters（过滤器），Proxy Server（代理服务器），Domain Name Server（域名服务器），NAT（网络地址转换），Safe Mail（安全邮件），Hardening（加固安全），VPN（虚拟专有网络），LOG（记录）等等。从安全性考虑，采用防火墙和代理服务器两道隔离屏障，内部网络区使用不能被 Internet 直接访问的虚拟 IP 地址，它可以通过路由器方便的浏览外部 Internet 网信息，但外部非法用户却无法轻易进入，减少了网络黑客入侵的威胁。

防火墙的实现从层次上大体上可以分两种：报文过滤和应用层网关。

报文过滤是在 IP 层实现的，只用路由器就可以完成，它根据报文的源 IP 地址、目的 IP 地址、源端口、目的端口及报文传递方向等报头信息来判断是否允许报文通过。报文过滤器的应用非常广泛，因为 CPU 用来处理报文过滤的时间可以忽略不计，而且这种防护措施对用户透明，合法用户在进出网络时，根本感觉不到它的存在，使用起来很方便。报文过滤的弱点主要是不能在用户级别上进行过滤，即不能识别不同的用户和防止 IP 地址的盗用，如果攻击者把自己主机的 IP 地址设成一个合法主机的 IP 地址，就可以很轻易地通过报文过滤器。

报文过滤存在的弱点可以用应用层网关解决，在应用层实现防火墙，方式多种多样，下面是几种应用层防火墙的设计实现。

①应用代理服务器

在网络应用层提供授权检查及代理服务。当外部某台主机试图访问受保护网络时，必须先在防火墙上经过身份认证，然后防火墙把外部主机与内部主机连接，它可以限制用户访问的主机、访问时间及访问的方式。同样，受保护网络内部用户访问外部网时也需先登录到防火墙上，通过验证后，才可访问。