

第 5 章 Internet 应用技术

丰富的应用是 Internet 能够流行的重要保证。本章将系统地讨论 Internet 应用技术发展的三个阶段以及 Internet 所提供的基本服务,并重点讨论基于 Web 的网络应用与基于 P2P 的网络应用。

5.1 Internet 应用发展分析

5.1.1 Internet 应用技术发展阶段

图 5-1 描述了 Internet 应用技术发展阶段示意图。从图中可以看出,Internet 应用技术的发展大致可以分为三个阶段。

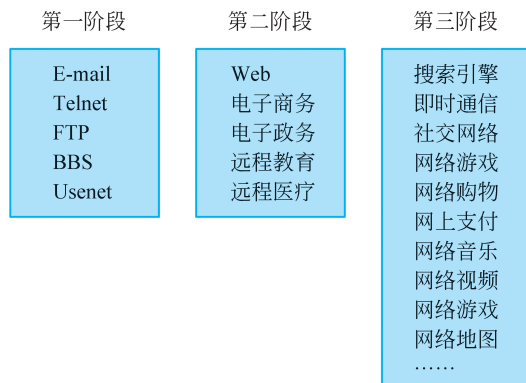


图 5-1 Internet 应用发展趋势

第一阶段互联网应用的主要特点:提供远程登录(TELNET)、电子邮件(E-mail)、文件传输(FTP)、电子公告牌(BBS)、网络新闻组(Usenet)等基本服务。

第二阶段互联网应用的主要特点:Web 技术的出现,以及基于 Web 技术的电子政务、电子商务、远程医疗、远程教育应用的快速发展。

第三阶段互联网应用的主要特点:各种新的互联网应用(例如,搜索引擎、即时通信、社交网络、网络购物、网上支付、网络音乐、网络视频、网络游戏、网络地图等)风起云涌,移动互联网将互联网应用推向一个新的高潮,物联网应用开始出现。互联网、移动互联网与物联网的应用成为新的经济增长点。

5.1.2 我国 Internet 的发展状况

根据中国互联网络信息中心(CNNIC)发布的《中国互联网络发展状况统计报告》的数据,2000—2022 年我国网民规模增长的趋势如图 5-2 所示。从 1994 年 4 月 20 日我国通过一条 64kb/s 国际专线实现与 Internet 连接,成为第 77 个接入 Internet 的国家之日算起,经过 27 年的发展,2022 年 12 月我国网民规模达到 10.67 亿,网民数量居世界第一;普及率达到 75.6%,超过世界平均普及率。

我国政府高度重视 Internet 应用对国民经济与社会发展的重要作用。在 2015 年发布的《国务院关于积极推进“互联网+”行动的指导意见》中指出:“互联网+”是把互联网的创新成果与经济社会各领域深度融合,推动技术进步、效率提升和组织变革,提升实体经济创新力和生产力,形成更广泛的以互联网为基础设施和创新要素的经济社会发展新形态。在全球新一轮科技革命和产业变革中,互联网与各领域的融合发展具有广阔前景和无限潜力,已成为不可阻挡的时代潮流,正对各国经济社会发展产生着战略性和全局性的影响。

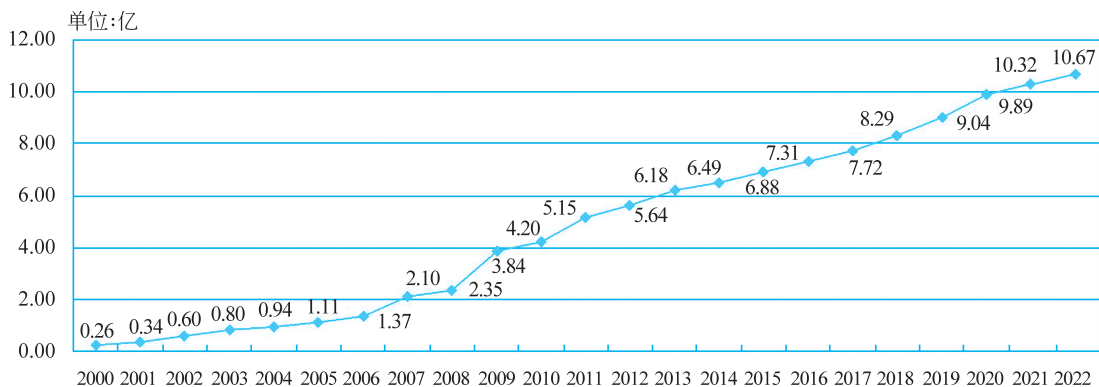


图 5-2 2000—2022 年我国网民规模增长趋势图

近年来,我国政府一直致力于推进“互联网+”发展,重塑创新体系、激发创新活力、培育新业态和创新公共服务模式,打造数字经济新优势。我国推进“互联网+”行动的基本原则如下。

(1) 坚持开放共享。营造开放包容的发展环境,将互联网作为生产生活要素共享的重要平台,最大限度优化资源配置,加快形成以开放、共享为特征的经济社会运行新模式。

(2) 坚持融合创新。鼓励传统产业树立互联网思维,积极与“互联网+”相结合。推动互联网向经济社会各领域加速渗透,以融合促创新,最大程度汇聚各类市场要素的创新力量,推动融合性新兴产业成为经济发展新动力和新支柱。

(3) 坚持变革转型。充分发挥互联网在促进产业升级以及信息化和工业化深度融合中的平台作用,引导要素资源向实体经济集聚,推动生产方式和发展模式变革。创新网络化公共服务模式,大幅提升公共服务能力。

(4) 坚持引领跨越。巩固提升我国互联网发展优势,加强重点领域前瞻性布局,以互联网融合创新为突破口,培育壮大新兴产业,引领新一轮科技革命和产业变革,实现跨越式发展。

(5) 坚持安全有序。完善互联网融合标准规范和法律法规,增强安全意识,强化安全管理和防护,保障网络安全。建立科学有效的市场监管方式,促进市场有序发展,保护公平竞争,防

止形成行业垄断和市场壁垒。

5.2 Internet 的域名机制

5.2.1 域名的概念

IP 地址解决了 Internet 的全局地址问题,通过 IP 地址可以找到唯一的一台主机。就像日常生活中使用的电话号码一样,IP 地址也是由一连串数字组成,例如“202.113.19.122”,人们通常难以记住这些数字。相对于 IP 地址,人们更喜欢通过名字来表示一台主机,例如“www.nankai.edu.cn”,每个字符都代表一定的含义,并且在书写上有一定的规律。这样,用户就容易理解,同时也容易记忆。因此,Internet 采纳了这种命名机制,这就是人们常说的域名机制。

如果 Internet 主机之间需要通信,在发送与接收数据时必须使用 IP 地址。尽管人们可以用名字来表示一台主机,但是在向这台主机发送数据之前,需要将它的名字转换为对应的 IP 地址。Internet 提供了将主机名转换为 IP 地址的服务。域名(domain name)就是主机的名字。域名系统(Domain Name System,DNS)是将域名转换为 IP 地址的服务。当应用程序需要处理一个域名时,它利用 DNS 将该域名转换为 IP 地址,并在接下来的通信中使用得到的 IP 地址。

在 Internet 的早期阶段,采用的是集中式的主机域名机制。网络信息中心(Network Information Center,NIC)维护一个主机文件(hosts.txt),其中保存主机域名与 IP 地址的映射表。早期的主机都是通过广域网接入 Internet,后来个人计算机开始大规模应用时,这些计算机通常是通过局域网接入 Internet。如果仍使用 hosts.txt 进行域名解析,提供域名服务的主机难以承载通信负荷。针对 Internet 主机数量剧增的情况,人们提出将域名系统划分为多个域,通过分布式的域名服务器提供域名服务。

图 5-3 显示了域名服务的工作过程。如果源主机想访问域名为“www.nankai.edu.cn”的目的主机,首先向本地网络中的 DNS 服务器(称为本地服务器)发送查询请求。如果本地服务器查到该域名对应的 IP 地址,向源主机返回包含该 IP 地址的响应;否则,本地服务器向上级 DNS 服务器发送查询请求。根据所处的位置和所起的作用,域名服务器可分为 4 种类型:本地服务器、权限服务器、顶级服务器与根服务器。在源主机获得目的主机的 IP 地址后,后续的

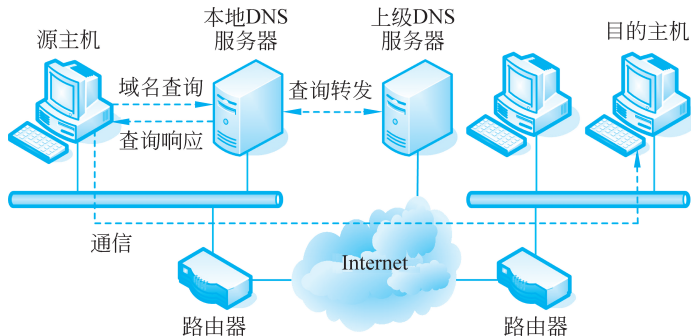


图 5-3 域名服务的工作过程

通信过程中将会使用该 IP 地址。因此,域名服务的核心技术是借助分层的 DNS 服务器结构来完成查询。

5.2.2 Internet 的域名结构

域名结构是由 TCP/IP 中的 DNS 来定义的。人们通常喜欢用简短的名字来命名计算机,但在 Internet 中必须使用长的名字来命名,以避免两台计算机采用相同的名字。为了保证计算机名在 Internet 中的唯一性,域名结构采用的是一种常见的思想:在每个名字后面添加额外的字符串(即后缀)。因此,计算机名通常包括 3 个部分:本机名、组织名与组织类型。例如,对于某个公司的 Web 服务器,可以被命名为“www.nankai.com”。

域名系统采用的是典型的层次结构,它将整个 Internet 划分为多个顶级域,并为每个域规定通用的顶级域名。表 5-1 给出了顶级域名分配方法。由于美国是 Internet 的发源地,因此其顶级域名是以组织模式来划分。例如,“com”表示商业公司,“edu”表示教育机构,“gov”表示政府部门。其他国家的顶级域名是以地理模式来划分,每个申请接入 Internet 的国家都作为一个顶级域出现。例如,“cn”表示中国,“fr”表示法国,“uk”表示英国,“jp”表示日本,“au”表示澳大利亚。

表 5-1 顶级域名分配

顶级域名	域名类型
com	商业组织
edu	教育机构
gov	政府部门
int	国际组织
mil	军事部门
net	网络中心
org	非营利性组织
国家代码	各个国家

NIC 将顶级域的管理权限授予指定的管理机构,各个管理机构再为自己管理的顶级域分配二级域,并将二级域的管理权限授予下属机构,这样就形成了域名系统的层次结构。图 5-4 给出了域名系统的层次结构。例如,教育管理机构拥有“edu”域的管理权限,它可以为下属大学分配各自的二级域,各个大学再为下属学院或系分配三级域。域名系统采用层次结构的最大优点是:各个组织在域中可以自由选择域名,只要保证在域中的唯一性,而不用担心与其他域的域名冲突。

5.2.3 我国的域名结构

CNNIC 负责管理我国的顶级域“cn”,将该域划分为多个二级域(如表 5-2 所示)。我国的二级域划分采用了两种划分模式:组织模式与地理模式。其中,前七个域对应于组织模式,而行政区代码对应于地理模式。按组织模式划分的二级域名中,“ac”表示科研机构,“com”表示商业组织,“edu”表示教育机构,“gov”表示政府部门,“int”表示国际组织,“net”表示网络中

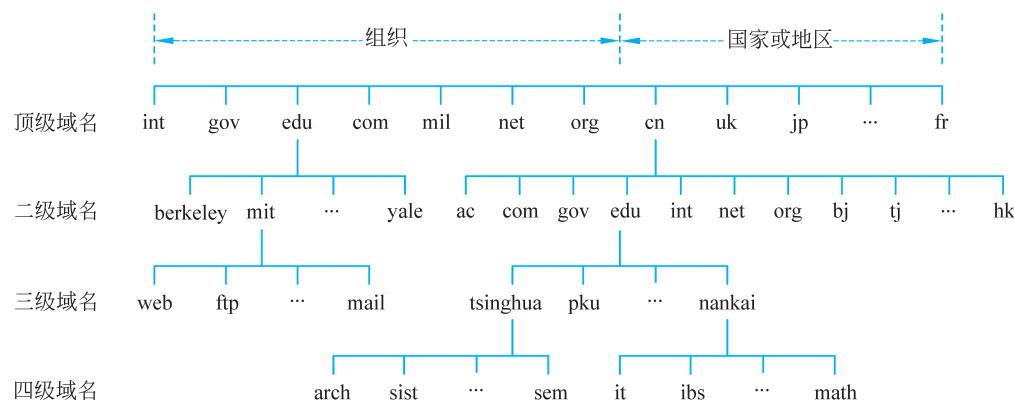


图 5-4 域名系统的层次结构

心,“org”表示非营利性组织。按地理模式划分的二级域名中,“bj”代表北京市,“sh”代表上海市,“tj”代表天津市,“he”代表河北省,“hl”代表黑龙江省,“nm”代表内蒙古自治区,“hk”代表香港地区。

表 5-2 二级域名分配

二 级 域 名	域 名 类 型
ac	科研机构
com	商业组织
edu	教育机构
gov	政府部门
int	国际组织
net	网络支持中心
org	各种非营利性组织
行政区代码	我国的各个行政区

CNNIC 将二级域的管理权限授予指定的管理机构,各个机构再为自己管理的二级域分配三级域。例如,CERNET 网络中心拥有“edu”域的管理权限,它可以为下属大学分配各自的三级域,各个大学再为下属学院或系分配四级域。域名的排列原则为:低层域名在前,其所属的高层域名在后,中间用符号“.”分开。域名的基本格式为:四级域名.三级域名.二级域名.顶级域名。图 5-5 给出了域名的例子。例如,“www.nankai.edu.cn”是南开大学的网站,“www.cnnic.net.cn”是 CNNIC 的网站。

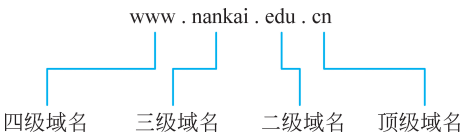


图 5-5 Internet 域名的例子

在域名系统中,每个域是由不同组织管理,这些组织将其子域交给其他组织管理。这种层次结构的优点是:各个组织在内部可自由选择域名,只要保证组织内的唯一性,而不用担心与其他组织的域名冲突。例如,南开大学是一个教育机构,则其主机域名为“nankai.edu”;如果一家公司也想用“nankai”来命名其主机,由于该公司是一个商业机构,因此其主机域名为“nankai.com”。在 Internet 的域名结构中,“nankai.edu”与“nankai.com”两个域名是相互独

立的。

5.3 Internet 的基本应用

5.3.1 电子邮件服务

电子邮件是 Internet 中最早提供、最受欢迎的服务之一,每时每刻都有数以亿计的人使用电子邮件进行通信。

1. 电子邮件的产生背景

电子邮件服务又称为 E-mail 服务,是指用户通过 Internet 收发电子形式的邮件。电子邮件是一种非常方便、快速和廉价的通信手段,这些都是电子邮件的基本特点。在传统通信中需要几天完成的投递过程,电子邮件仅用几分钟、甚至几秒就能完成。目前,电子邮件已成为网络用户的常用通信手段之一。早期的电子邮件只能传输文本信息,当前还可以传输 HTML 格式信息。

电子邮件是伴随着 Internet 而发展起来的。1971 年,电子邮件诞生于美国马萨诸塞州的 BBN 公司,该公司受聘于美国军方参与 ARPANET 的建设。电子邮件的发明者是 BBN 公司的 Ray Tomlinson,他在已有的文件传输程序的基础上,开发了在 ARPANET 中收发信息的邮件程序。为了让人们拥有易于识别的邮件地址,他决定用“@”隔开用户名与邮件服务器地址,这就是现在使用的电子邮件地址的起源。

由于最初的 ARPANET 中的结点数很少,当时并没有多少人使用电子邮件,这种情况直到 ARPANET 转向 Internet 才得到改变。最初,电子邮件受到网络传输速度的限制,那时用户只能发送一些简短的信息,无法像现在这样发送多媒体信息。1988 年,第一个图形界面的邮件客户机软件问世,它就是著名的 Euroda 软件。后来,Netscape 与 Microsoft 公司相继推出邮件客户机软件。随着 Internet 用户数量的急剧增加,电子邮件逐渐成为一种流行的 Internet 服务。

2. 电子邮件的概念

我们首先分析现实社会中的传统邮政系统。现实中的邮政系统已有近千年的历史。每个国家负责管理自己国内的邮政系统,按照省、市、区(县)建立不同级别的邮政系统,底层的邮政系统在自己管辖范围内设立邮局,邮局在单位或个人的家门口设立邮箱,邮递员完成邮件的接收、转发与投递工作。邮政部门需制定相应的通信协议与管理制度,甚至需要规定信封按什么规则来书写。由于有整套严密的组织结构与通信规程,因此能保证邮件及时、准确地送到目的地。

电子邮件系统与现实中的邮政系统有相似结构。两者之间的不同点主要在于:邮政系统是由人工控制各种运输设备来运转,电子邮件是在 Internet 中通过计算机、应用软件与协议来运转。电子邮件系统中同样需要邮局与邮箱,它们是邮件服务器(mail server)与电子邮箱(mail box)。其中,邮件服务器负责发送与接收电子邮件,电子邮箱负责存储电子邮件。另外,需要规定电子邮件的书写格式与传输协议。

邮件服务器是整个电子邮件系统的核心。邮件服务器的主要功能包括:接收发件人通过客户机软件发送电子邮件,并按收件人地址转发给对方的邮件服务器;接收其他邮件服务器发

送的电子邮件,并按收件人地址存储在相应的邮箱;根据收件人的要求将电子邮件发送给收件人的客户机软件。电子邮箱由提供电子邮件服务的机构来建立,通常被称为电子邮件账号(包括邮件地址与密码)。

电子邮件地址(E-mail Address)是邮件服务器中的邮箱地址。电子邮件地址是由汤姆林森最早提出的,使用“@”隔开用户名与邮件服务器地址,这是现在使用的电子邮件地址的起源。邮件地址的关键是保证每个地址的唯一性,以便邮件经过邮件服务器的转发,并被准确投递到相应的邮箱中。电子邮件地址的具体格式为:用户名@主机名。其中,用户名是用户在邮件服务器中的邮箱名,主机名是邮箱所在的邮件服务器名。图 5-6 给出了电子邮件地址格式。例如,电子邮件地址为“island@nankai.edu.cn”。其中,“island”为邮箱名,“nankai.edu.cn”为邮件服务器名。



图 5-6 电子邮件地址格式

3. 电子邮件的工作原理

电子邮件服务核心技术主要包括:简单邮件传输协议(Simple Mail Transfer Protocol, SMTP)、邮局协议(Post Office Protocol, POP)和交互式邮件存取协议(Interactive Mail Access Protocol, IMAP)。这些协议是邮件服务使用的应用层协议,它用来实现客户机与邮件服务器之间的通信。其中,SMTP 是电子邮件发送协议,用来将邮件从客户机发送到邮件服务器;POP 与 IMAP 是电子邮件接收协议,用来将邮件从邮件服务器接收到客户机。

电子邮件服务采用客户机/服务器模式。电子邮件服务包括两个组成部分:邮件客户机与邮件服务器。其中,邮件客户机是电子邮件服务的使用者,邮件服务器是电子邮件服务的提供者。图 5-7 给出了电子邮件系统结构。电子邮件的发送与接收采用不同的协议,客户机与邮件服务器都需要实现两种协议,SMTP 与 POP 或 IMAP 中的一种。邮件客户机通常包括两个部分:SMTP 客户与 POP 客户。邮件服务器通常包括 3 个部分:邮箱、SMTP 服务器与 POP 服务器。

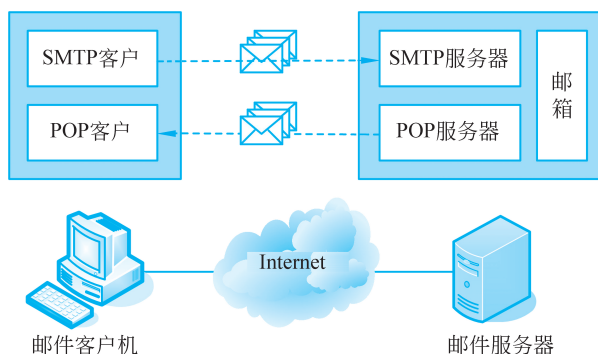


图 5-7 电子邮件系统结构

电子邮件被存储在邮件服务器的相应邮箱中。图 5-8 给出了电子邮件的工作原理。用户通过客户机与邮件服务器建立连接,并向邮件服务器发出邮件发送或接收请求;如果是用于邮件发送的 SMTP 请求,客户机将邮件发送到自己的邮件服务器,该服务器再将邮件发送到最终的邮件服务器,整个过程可能经过多个邮件服务器转发;如果是用于邮件接收的 POP 请求,邮件服务器从相应的邮箱中读取邮件,客户机接收返回的邮件后进行解释,并将邮件信息显示在客户机的屏幕上。

用户通过邮件客户机程序访问电子邮箱中的邮件。邮件客户机需要安装在用户的计算机或移动终端(例如手机)中。用户通过邮件客户机登录进入自己的邮箱,这时能看到邮箱中所有邮件的列表,包括发件人地址、主题、时间、附件等信息,并决定对某封邮件进行哪种操作。邮件客户机的基本功能主要包括:书写与发送邮件;接收、转发、回复与删除邮件;邮箱与通讯簿管理。常见的邮件客户机程序包括 Microsoft Outlook Express 或 Outlook、Mozilla Thunderbird、Foxmail 等。

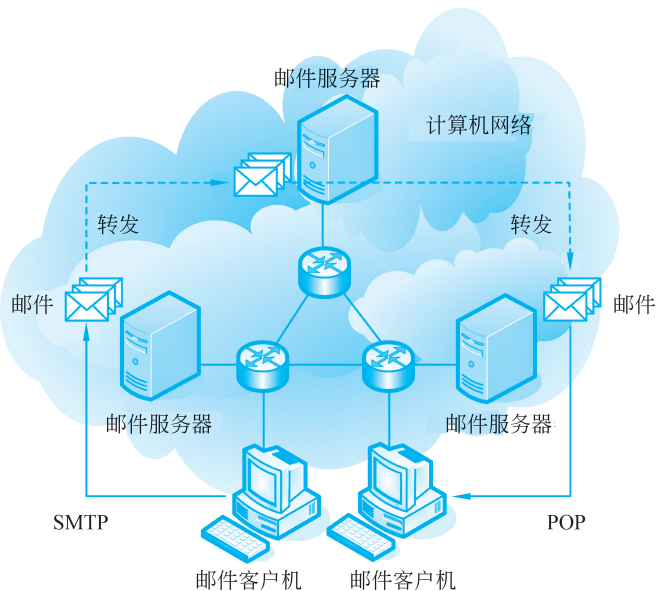


图 5-8 电子邮件的工作原理

4. 电子邮件的信件格式

电子邮件与普通的邮政信件相似,也有信件格式方面的统一规定,以保证邮件能在不同的邮件服务器之间转发。1982 年,RFC 822 文档定义邮件信件格式,它是目前电子邮件遵循的信件格式标准。2001 年,RFC 2822 文档定义最新版的信件格式,对早期的信件格式没有大幅更改。SMTP 将邮件整个封装在邮件对象中,其中所有信息由 ASCII 码组成。电子邮件可以由多个报文行组成,各行之间用回车(CR)与换行(LF)分隔。

图 5-9 给出了电子邮件信件格式。邮件对象包括两个部分:信封与邮件内容。实际上,信封就包含两种 SMTP 请求,用来给出收件人与发件人地址。电子邮件包括两个部分:邮件头(mail header)与邮件体(mail body)。其中,邮件头由邮件的相关信息构成,其中的部分信息由系统自动生成,例如发信人(From)、时间(Date)等;其他信息由发件人自行输入,例如收信人(To)、主题(Subject)与抄送人(Cc)等。邮件体是指需要发送的邮件正文部分。

SMTP 只能传输由 ASCII 组成的邮件信息,这样无法支持非 ASCII 码的语种,例如中文、法文、德文与俄文等。另外,它不支持邮件中附带的二进制文件,例如图片、音频、视频、可执行文件等。1993 年,多用途 Internet 邮件扩展(Multi-purpose Internet Mail Extensions,MIME)出现,它是一种辅助性的邮件编码协议。MIME 可通过 SMTP 来传输非 ASCII 码的数据,这样使电子邮件的用途变得更广泛。

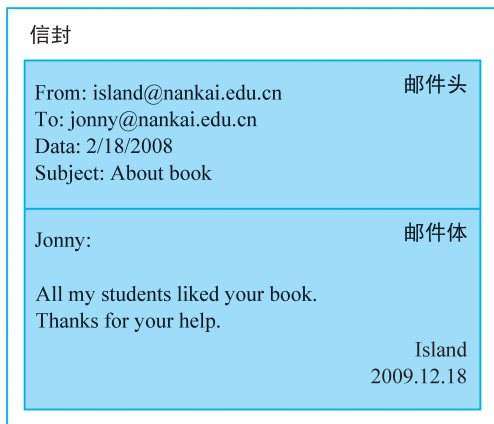


图 5-9 电子邮件信件格式

5.3.2 文件传输服务

文件传输是 Internet 中最早提供、最受欢迎的服务之一,每时每刻都有数以亿计的文件通过 Internet 传输。

1. 文件传输的产生背景

文件传输服务采用文件传输协议(File Transfer Protocol,FTP),因此它又被称为 FTP 服务。文件传输允许用户在两台计算机之间传输文件,并保证文件在 Internet 中传输的可靠性。FTP 服务伴随着 Internet 发展起来,它是 Internet 最早提供的服务之一。1971 年,第一个 FTP 技术文档出现标志着 FTP 诞生。在 Web 服务出现之前,人们主要通过 FTP 服务来共享文件资源。

FTP 服务器是指提供 FTP 服务的主机,它可看作一个容量很大的文件仓库。图 5-10 给出 FTP 目录结构的例子。FTP 服务器中以目录结构来保存文件,用户需逐级打开目录找到文件,然后传输其中某个文件。FTP 服务的目录结构带来使用上的不便。1990 年,第一个 FTP 搜索引擎 Archie 出现,它被认为是现代搜索引擎的鼻祖。在 Internet 发展初期,FTP 服务通信量占整个网络的三分之一。直到 1995 年,Web 服务的通信量开始超过 FTP 服务。

2. 文件传输的基本概念

FTP 服务采用客户机/服务器模式。FTP 服务包括两个组成部分:FTP 客户机与 FTP 服务器。其中,FTP 服务器是提供 FTP 服务的计算机;FTP 客户机是用户的本地计算机。图 5-11 给出了下载与上载的概念。下载是将文件从 FTP 服务器传输到客户机,上载是将文件从 FTP 客户机传输到服务器。如果用户使用 FTP 服务传输文件,并不需要对文件进行复杂的转换工作,因此这种服务的工作效率很高。

如果用户要使用 FTP 服务器提供的服务,首先从 FTP 客户机登录到 FTP 服务器,这时需要输入 FTP 服务器名与账号。每个 FTP 服务器都有自己的服务器名,并保证 FTP 服务器名在全球范围内的唯一性。例如,FTP 服务器名为“ftp.pku.edu.cn”。其中,“ftp”表示提供 FTP 服务,“pku.edu.cn”表示北京大学的主机。当用户登录到 FTP 服务器时,通常需要提供用户账号(包括用户名与密码)。有些 FTP 服务器是专用的,只允许拥有合法账号的用户使用。

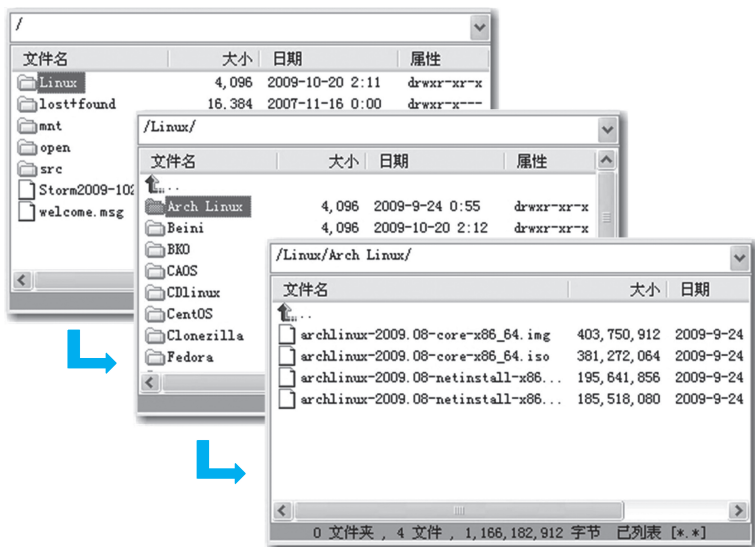


图 5-10 FTP 目录结构的例子

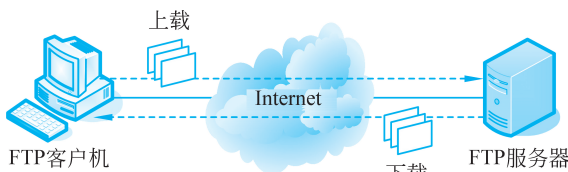


图 5-11 下载与上传的概念

目前,很多 FTP 服务器可提供匿名 FTP 服务。这类服务器没有为每个用户设置专用账号,它会提供一个公开的用户账号(通常为 anonymous),并赋予该账号访问 FTP 服务器中公共目录权限。如果用户需要访问匿名 FTP 服务器,可以用“anonymous”作为自己的用户账号,访问位于公共目录下的所有文件。为了保证 FTP 服务器自身的安全,多数的匿名 FTP 服务器只提供文件下载服务。

3. 文件传输的工作原理

FTP 服务使用的应用层协议是 FTP。FTP 服务需要建立两个连接:控制连接与数据连接。其中,控制连接用来传输 FTP 命令与响应,数据连接用来传输实际的文件数据。图 5-12 给出了 FTP 服务的工作原理。用户通过客户机与 FTP 服务器建立控制连接,向 FTP 服务器发出登录请求并提供账号,然后可建立数据连接并完成文件传输。控制连接在数据连接之前建立,并且控制连接在数据连接之后释放。

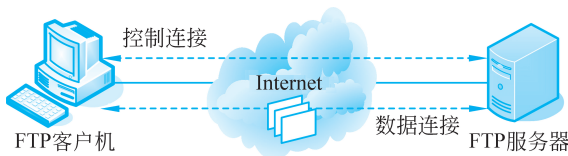


图 5-12 FTP 服务的工作原理