普通高等教育“十一五”规划教材

Windows 网络编程技术

胡 鸣 主编

科学出版社
www.sciencep.com

·普通高等教育“十一五”规划教材·

Windows 网络编程技术

胡 鸣 主编

科学出版社

北京

内 容 简 介

本书通过应用实例由浅入深引入 VC++ 集成开发环境下的网络编程技术。全书突出网络编程技术应用背景，发掘编程技术与网络应用开发的逻辑联系。内容主要包括 Windows 网络编程基础、单机资源共享的应用编程、基于 NetBIOS 网络编程、基于 WinSock 网络编程、直接网络编程和高级网络编程等。

本书有光盘源码，适合那些具有一定 C 和 C++ 语言基础，期望逐步加强网络编程能力或网络编程相关综合训练的计算机类专业以及网络应用相关专业学生和技术人员使用。教师可以根据课时要求选择讲授本书的章节。

图书在版编目（CIP）数据

Windows 网络编程技术/胡鸣主编.—北京：科学出版社，2008
普通高等教育“十一五”规划教材
ISBN 978-7-03-022286-2

Ⅰ.W… Ⅱ.胡… Ⅲ.窗口软件，Windows—程序设计—高等学校—教材 Ⅳ.TP316.7

中国版本图书馆 CIP 数据核字（2008）第 084746 号

责任编辑：张颖兵 吉正霞／责任校对：梅 莹
责任印制：董艳辉／封面设计：苏 波

科学出版社出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com
武汉中科兴业印务有限公司印刷
科学出版社发行 各地新华书店经销
*
2008 年 6 月第 一 版 开本：787×1092 1/16
2008 年 6 月第一次印刷 印张：15 1/4
印数：1—4 000 字数：345 000
定价：29.80 元（含光盘）
（如有印装质量问题，我社负责调换）

前　言

当你在电脑上浏览网页、上传/下载文件(包括音乐与电影)、收发电子邮件、进行网络聊天和置身网络游戏中时,你已经通过互联网服务把你的生活、娱乐和工作与网络关联起来,从而提高生活的品质、娱乐的层次和工作的效率。对于计算机类和网络类专业的学生与工程技术人员而言,显示其专业优势之一就是能够利用网络编程技术,开发各具特色的网络应用程序,为网络用户提供更为人性化的网络服务。本书作为一本自主工程化培养理论与实践结合的高级计算机编程人才的书籍,将引导你开发网络应用程序,感悟各种网络编程技术的意义与作用,逐步获得工程项目开发的意识。

网络编程通过选择开发环境——操作系统(Windows 和 Unix)和编程语言(C++和 Java),在网络应用需求的驱动下,开发网络应用程序。本书选用大家熟悉的 Windows 开发环境和使用广泛的且与 Windows 开发环境配套紧密的 Visual C++开发工具,从而有利于上机实验环境的配置、增强编程开发的灵活和提高网络应用程序运行的效率。

本书的读者对象

本书的读者最好进行了操作系统和计算机网络知识的学习,对应用程序如何共享系统资源、网络资源和信息的基本流程有了一定的认识;学习了数据结构的知识,对计算机编程的基本要素——数据结构和算法有了一定的了解;也经过了 C/C++编程的基本训练,掌握简单应用编程的思想。

当然对操作系统、计算机网络、数据结构以及 C 语言基础掌握得足够好,可以加快对 Windows 网络编程技术的理解与掌握。不过,本书并不追求读者对这些知识的全面理解和掌握,因为本书将在需要的时候,引入相关知识的概念,并在正文和附录中有所体现,以此来保证本书概念和引用的完备性。

本书的特点和内容

网络应用编程,对学生与开始涉猎的工程技术人员而言跨度较大,既涉及两个应用程序通过网络通信(包括应用层、运输层、网络层和数据链路层等不同层次的通信问题),也涉及单机提高效率的编程技术(异步 I/O 控制、多进程通信、多线程通信、动态链接库)以及相关的同步控制。相对于单机单线程的编程,网络应用编程综合性强,但对许多具体网络综合应用程序开发,网络编程又只是网络综合应用的基础。因此,网络编程无论对学生还是对教师而言都具有挑战性。如果经过网络编程的训练和学习,学生不能把握网络编程开发脉络,那么他们就会随着网络编程技术不断引入犹如迷宫探险,并伴随更多变量类型和概念的迷茫以及编程调试和运行挫折的打击,而对网络编程产生恐惧。这是教师与学生都不愿看到的。

本书的作者经过几年不同层次学生和不同学时(32 到 60 学时)网络编程教学实践和

理论的探索，为本书的编写积累了一套完整的素材并据此拟定了切实可行的内容组织计划，逐步提炼出本书编写的特色。本书结合附带的光盘逐步引导读者自主掌握 Windows 网络编程技术，具有以下特点：

(1) 设计或者选择代表性强、可读性好、层次分明以及复杂度逐步提高的程序实例。

(2) 突出网络编程技术应用背景，发掘编程技术与网络应用开发的逻辑联系。

(3) 优化内容组织以便读者完成基础部分向编程部分，单机应用向网络应用，简单网络编程向复杂网络编程，单一网络应用向综合网络应用以及应用需求到算法设计到程序代码实现的 5 个过渡。

(4) 为了突出网络编程开发的脉络，让读者有拨云见日的感受，将协议族框架、具体编程接口功能调用和参数的说明等放在附录部分。

本书经过内容组织优化易于阅读，各章节内容组织如下：

第 1 章讲述 Windows 网络编程基础，从网络应用到 Windows 环境的编程条件到 Windows 网络编程，描述了应用驱动编程开发的机制。

第 2 章详细介绍单机资源共享的应用编程，涉及后面章节网络编程需要的单机资源共享编程技术，如多进程、多线程、同步控制和动态链接库。

第 3 章详细介绍基于 NetBIOS 网络编程，通过应用实例描述应用层的网络编程，涉及 NetBIOS 编程接口和数据报与会话两种通信机制。

第 4 章详细介绍基于 WinSock 网络编程，通过应用实例重点描述运输层的网络编程，涉及数据报、会话通信以及异步 I/O 的概念与编程技术。

第 5 章详细介绍直接网络编程，通过应用实例描述网络层和数据链路层的网络编程，涉及原始套接字和 Winpcap 数据包捕获等编程技术。

第 6 章详细介绍高级网络编程，通过应用实例描述综合了 MFC 框架的网络编程，涉及网络编程与 MFC 框架的简单组合、基于 MFC 的 Socket 类和 WinInet 的编程技术。

附录 1～5 分别介绍第 1 章到第 5 章需要的协议框架、功能调用的库函数和参数以及错误代码说明，并且作分类处理，以便在需要的时候引用它们。

结构上本书第 1 章为基础部分，它涉及编程的基本框架和开发平台的结合；第 2 到 6 章为编程部分。对于编程部分，第 2 章为单机应用编程部分，第 3 到 6 章为网络应用编程部分，而网络应用部分也综合了单机应用编程部分的编程技术。对于网络应用编程部分，第 3 到 5 章为单一界面的网络编程部分，第 6 章为基于 MFC 图形界面的网络编程部分；单一界面的网络编程应用都可以经过不同形式的变换得到具有图形界面的网络应用。从第 3 到 5 章，网络编程从应用层到运输层再到网络层和数据链路层，体现了简单网络编程向复杂网络编程的过渡。

此外，带 * 号的章节表示教师可以根据课时和授课对象选择是否讲授。第 3 章基于 NetBIOS编程接口，Windows XP 以上的操作系统自动关闭，而第 5 章直接网络编程部分涉及网络维护与网络安全方面。因此在学时低于 48 的情况下，教师可以根据学生专业情况，选择第 3 章或第 5 章不作为授课范围。

本书适合于计算机类和网络类专业高年级学生，他们经历了操作系统、数据结构、计算机网络和 C 语言的学习，并可以通过这门课程的编程训练和学习，综合所学的知识提高综合应用编程的能力，为工程项目设计与开发打下坚实的基础。本书也适合于那些需

要掌握 Windows 网络编程技术的工程技术人员，以便进一步提高网络综合应用编程与开发能力。因此，本书的目标就是全力打造高级的计算机编程人才，以解决 IT 业中这类人才紧缺的问题。

本书由胡鸣老师任主编，负责全书的总体规划、审稿与最后统稿，另外负责第 1 章、第 3 章和第 4 章的编写；有过网络编程教学与教改经验的聂刚老师编写了第 5 章；有过面向对象编程教学与教改经验的孔维广老师编写了第 6 章；有过单机资源共享应用编程经验的彭涛老师编写了第 2 章。

由于本书的编写工作较为仓促，编者的水平有限，书中不可避免存在疏漏和不足，敬请广大读者不吝赐教，我们会在适当时间进行修订和补充。读者有任何建议和批评都可以通过电子邮件发到 stereotype@263.net。我们的所有努力都是为选择最佳支点，以期最快地助推读者达到 IT 编程的最高境界！

编　者

2008 年 3 月

目　录

第 1 章　Windows 网络编程基础

没有人天生就会编写网络应用程序！对于初学者，网络应用软件（程序加数据）开发是一个过程，其核心就是网络应用程序的开发。它的第一步就是做好准备，了解引发网络编程开发的动力，Windows 系统环境下应用程序运行与产生的机制，进一步，Windows 网络应用程序开发的特点。不知道这些，就不了解 Windows 网络应用软件开发的机制，就没法形成自主开发工程项目的意识。本章将详细描述 Windows 网络应用需求驱动网络应用软件开发的机制，它是 Windows 网络编程的基础。

1.1　Windows 网络应用

了解 Windows 网络编程的目标、Windows 网络应用程序运行的环境以及 IT 市场上这类应用需求的程度和范围，构成 Windows 网络应用软件开发的动力。本节将通过对 Windows 网络应用的讨论与分析，理解 Windows 网络编程开发的意义。

1.1.1　常用的网络应用实例

当今，人们每天的生活几乎都离不开计算机。为何如此？因为计算机提供的窗口（Windows）不仅可以享受本地计算机系统资源提供的信息服务，而且还可以通过互联网享受全世界其他计算机系统资源提供的信息服务。个人需要互联网、企业需要互联网、学校需要互联网，政府需要互联网，乃至社会的方方面面都需要互联网，需要互联网提供的服务。因此，Windows 网络编程的目标，就是在 Windows 环境运行的计算机上开发能够运行的网络应用程序作为平台，为用户提供形形色色的互联网服务。

互联网提供的服务并不陌生。用户了解世界最快捷的方式就是万维网（WWW）服务。通过在一台计算机上运行浏览器（例如，Windows 操作系统捆绑的 IE），浏览另一台机器上通过网络应用程序提供的万维网文档，这些文档被称为网页，它们是两个应用程序处理的数据部分。这个万维网服务文如其名，功能强大，可以在线浏览新闻，提供文档搜索引擎，在线收发邮件，在线上传/下载文件，在线收看电影和收听音乐，构建各种信息窗口（个人主页空间、博客、在线论坛/BBS/讨论组、网上校友录等），获取信息（产品服务查询、工作信息查询、医疗健康服务查询、政府信息查询等），构建各种服务平台（网上购物、招聘、金融、杂志、教育、销售、预订和电子政务等）。

用户最常使用的文字通信是电子邮件服务。通过计算机上运行的用户代理程序（Windows 捆绑的 Outlook 或者单独安装的 Foxmail 等），借助收发双方类似邮局的计算机邮件服务程序的支持，实现用户非即时（双方可以不在场）的电子邮件通信。电子邮件服务与万维网服务相比，针对性强，效率更高。不过，不同用户代理程序（如 Outlook 和 Foxmail）的存储邮件数据的格式不同。

用户共享计算机软件资源的方式是文件传输服务。文件传输服务的实现，首先要求

储存用于传输文件的服务器上的服务程序能够支持文件传输功能的实现。用户需要安装专用文件传输应用程序，如 CuteFTP，BpFTP，LeapFTP 和 FlashFTP 等，或者与万维网配合使用安装网际快车 Flashget、网络蚂蚁 NetAnts、迅雷和 BT 等应用程序实现高效率文件下载。不同的应用程序处理中间未完成数据的方式可以不同。

用户与其他人最好的交流方式，就是个人聊天服务，或称个人即时通服务。目前有各种各样的聊天服务程序，如腾讯网站的 QQ、网易网站的 POPO 与微软 MSN 网站的 MSN。只有下载安装了同类聊天软件的用户才能互通。这些用户通过加入好友，与对方进行文字、语音和视频聊天或者互传文件和邮件，进入或自创专题聊天室和群，有的甚至可以进入网络游戏室。不同聊天软件提供不同风格的功能，它们处理用户传递的数据格式也不同，这就是为什么 QQ 传送的文件有时不能按原来的文件格式打开的原因。

用户最能展示自己的娱乐方式，就是毁誉参半的网络游戏服务。除了聊天软件附带的和游戏网站提供的普通网络游戏之外，最具挑战与震撼作用的就是那些用户可以进入角色的、全新体验的、类似电视剧的网络游戏，如文明、三国、魔兽世界、梦幻西游等。用户需要下载所选的网络游戏安装软件，然后借助管理游戏计算机应用程序的支持，进入游戏的角色，伴随剧情的演绎和动态效果获得在线娱乐体验。这与电视剧的不可交互不同，游戏用户的参与会改变剧情的发展。因此，不同于提供通用功能服务的网络应用软件，这类软件注重艺术的创新和在线娱乐参与，并随着用户全程体验的结束，从用户的主机应用程序中消失。因此，开发这类网络应用程序的潜力较大。

用户最需要解决信息服务麻烦的，就是维护网络安全服务。可以通过安装抗病毒软件与防火墙解决这类问题，如 WebWasher，Antivir，AVK，Symantec，Kaspersky，F-Secure和 Rising(瑞星)等，防止网络病毒、弹出式广告、网络入侵与攻击(包括木马)、网上收费陷阱与虚假信息、垃圾邮件、诱骗/欺诈/网络钓鱼、网上不良信息以及隐私泄露等。这是一场猫和老鼠的博弈。由于网上攻击者和病毒制造者不断改变攻击方式或提供新的病毒变种，提供网络安全软件的公司几乎需要每天更新用户的安全数据库。因此，用户计算机的安全应用程序需与提供该软件的公司计算机上的网络应用程序通信，以保证网络安全软件的更新。

1.1.2 网络应用的特点与运行环境

网络应用软件开发者，仅仅停留在网络服务的用户上是不够的，还需要了解网络应用程序运行的大环境。上面的网络应用实例显示，网络服务是通过服务方计算机运行的网络应用程序与用户计算机上安装的网络应用程序的协作处理某种格式数据来实现的。一般它们有以下几个方面的特点：

(1) 网络服务至少需要两台以上计算机运行应用进程通信的支持。

(2) 每台计算机运行的网络应用进程与该机器上的其他应用进程共享所在机器上的系统资源。

(3) 网络应用进程之间的通信由标准化的 I/O 接口(网卡或内/外置 modem，视本地网络不同而不同)以及连接双方网络硬件资源的支持。

(4) 网络应用进程的通信，需要区分不同的机器以及区分同一机器上不同的网络应用进程。网络服务功能需要网络寻址和数据转发机制，它们由参与通信的计算机网络协

议软件(如 TCP/IP 协议族)支持。

(5) 实现广泛的功能扩展,即具有虚拟设备(把网络上任何一台其他计算机看做使用接口的计算机设备)功能,需要网络通信的进程能够正确解析对方的信息,因此它们必须建立标准的通信规范。

上面特点显示,网络应用程序运行环境复杂,不仅需要共享通信双方计算机系统资源,而且需要共享网络上相关网络设备的系统资源。图 1-1 给出网络应用进程通信的运行环境,左边为经过路由器连接不同网络的网络拓扑图,右边为从网络应用程序看到的互联网络。

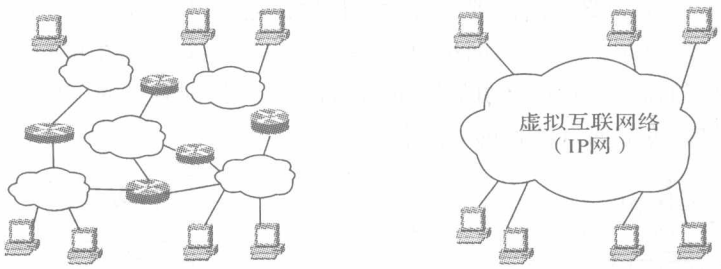

图 1-1　网络应用程序运行的环境

是什么使得网络应用程序看到的是图 1-1 右边的虚拟互联网络? 这就是各个计算机和网络设备上软件支持的互联网体系结构,目前最为流行的是 TCP/IP 协议族,如图 1-2 所示。TCP/IP 协议族的各层功能定义参见附录 1.1。TCP/IP 协议族只定义了应用层、运输层和网络层的协议,而 ISO/OSI 参考模型还定义了数据链路层和物理层。对 TCP/IP 协议族而言,其属于底层网络,一般称之为网络接口层。不同的底层网络有不同的标准,不属于 TCP/IP 协议族定义的范畴。对于网络应用程序而言,网络中间路由器需要网络层支持,转发连接不同底层网络的数据,因此图 1-1 将左边网络拓扑简化为右边的 IP 网络,因为网络层的核心是 IP 协议。

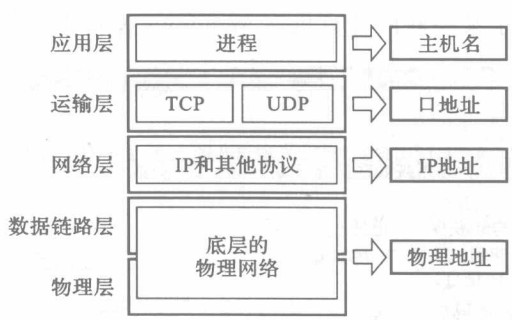

图 1-2　TCP/IP 协议族的结构功能

网络应用程序数据传输,需要机器在底层网络中的物理地址,即网络接口卡地址,在底层网络内寻址;跨不同底层网络传输需要机器在虚拟网络的逻辑地址,即 IP 地址,在网络层寻址;机器上识别不同应用进程的数据需要进程的接口地址,即端口地址;应用程序提供用户服务时需要提供网络服务的主机名。常用服务提供方的网络应用进程使用公认(或称保留)端口号(全局分配 0～1023 之间),而服务使用方使用自由(或称急用)端口号(本地分配:注册的 1024～49151 和动态绑定的 49152～65535)。任何主机上与网络应用

相关的程序都具有自己的端口号，同一主机上不同的网络应用程序或使用不同的网络功能，将使用不同的端口号。对于使用主机名跨越网络访问其他系统时，应用进程需要将主机名转换到 IP 地址（这一般需要域名服务 DNS），应用进程数据穿过不同底层网络传输时也需要 IP 地址与物理地址的转换（地址解析如 ARP）。

图 1-3 给出了两个处于不同计算机上应用进程数据传输的模型，其中，m，A 和 j 分别为数据发送方网络接口卡地址、IP 地址和应用进程的端口号；n，P 和 k 分别为数据接收方的网络接口卡地址、IP 地址和应用进程的端口号；x 为起始路由器在发送方机器底层网络中接口卡的地址，y 为终止路由器在接收方机器底层网络中的接口卡地址，如果数据发方机器和收方机器位于同一底层网络，则 x＝n，y＝m。除了发方和收方负载的影响之外，随着参与通信的底层网络、路由器数量、它们负载和处理容量的不同，应用发送进程和应用接收进程数据传输的效率也随着变化。

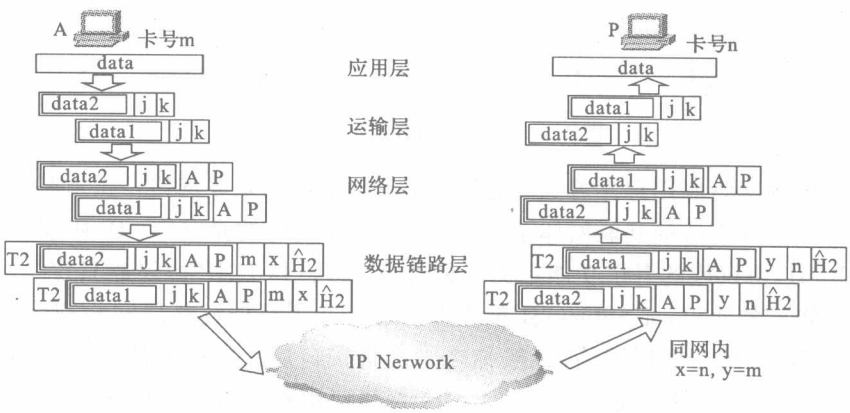

图 1-3　网络应用程序数据流的传输模型

要获得网络环境的支持，除了安装相关的网络接口驱动程序之外，还需要配置网络必需的参数。图 1-4 给出了某一联网计算机在 Windows 操作系统中的配置界面。

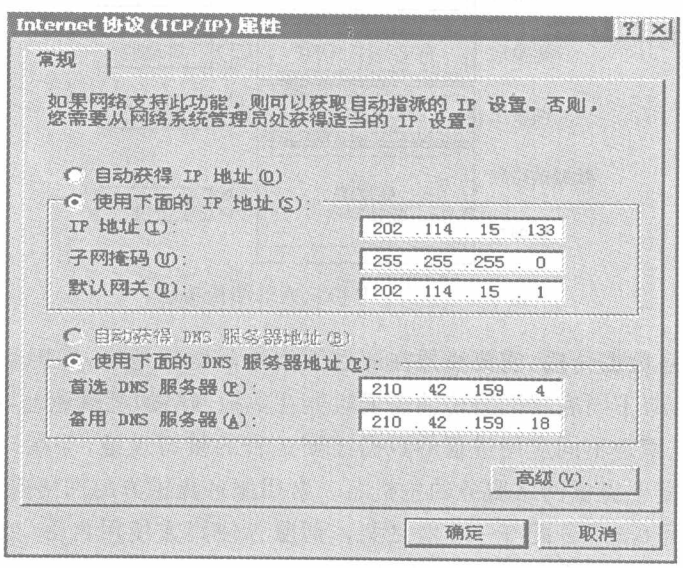

图 1-4　网络应用程序运行计算机的网络环境配置

对于自动获得 IP 地址，网络提供商通过 DHCP 协议和提供此类服务的计算机来实现；不能自动获得 IP 地址，则像图 1-4 中那样人工配置网络参数，如 IP 地址、子网掩码、默认网关、首选 DNS 服务器和备用 DNS 服务器。如果一个计算机有多个网络接口，则针对连接不同网络的每个接口，根据网络提供商的要求配置不同的网络参数。网络环境的保障，是计算机上的网络应用程序正常运行的基础，同时也体现它们对网络基础设施的依赖。

1.1.3 网络应用前景

如果读者对典型网络应用实例和网络应用运行环境展开进一步思考，就会产生网络编程开发的危机感。这么多网络应用软件进入我们的生活，广泛流行而成熟的 TCP/IP 协议族的软件对用户计算机的支持以及网络应用程序面临复杂多变的网络运行环境！网络编程开发的初学者怎么能够在 IT 市场上找到自己的位置？所编写的网络应用程序怎么能驾驭这些庞杂善变的网络环境？我们的回答只有一个：精英需要打造！从已有的网络应用实例和网络编程技术的应用中获取开发能力，从网络应用需求的综合分析中获得网络应用软件的市场定位，从用户需求分析的拓展思维中获取驾驭网络环境的灵感，从而提出网络技术应用的优化策略。下面从常用的网络应用实例和现有网络基础设施的更新，看网络应用的前景。

对于万维网服务，用户的 IE 浏览器提供用户可切换的、通过网络看世界的窗口。窗口切换的能力依赖于 Web 网（将各种信息源关联起来），其能力高低反映 Web 定位 URL（或称扫描）到有用窗口的精度和速度。一旦定位到某个窗口（由特定网站应用进程提供），就涉及窗口交互的信息和专业性服务；每个窗口可以同时为多个用户服务（点击率），这涉及可用性、及时性、稳定性和可靠性等，不同的窗口，服务要求、目标和功能也不同。用户计算机上浏览器可能限定到 IE，但用户可以浏览成千上万独具特色的窗口，每个窗口的服务方提供计算机应用程序各自实现自己的目标、业务和功能。要提高 Web 能力，窗口信息、功能以及服务描述都需要知识化形成语义网，而机器解释领域知识需要特定应用程序提供的服务支持。因此这类应用的市场潜力较大，这类应用涉及基于平台（Unix 和 Windows 平台的 C/C++和数据库开发）、跨平台（Java 与数据库开发）、Web（javascript，CSS/HTML/XHTML，flash action script，XML/XSLT）等产品的开发。基于平台的网络应用程序重点在于性能优化。

通信双方不必同时在线的电子邮件服务，一般在 3（邮件双方在同一个邮件域名中）到 4 个计算机应用程序（不必同时运行）支持下完成。用户计算机上的用户代理和提供邮件服务计算机上存储转发应用程序都有网络应用市场。任何单位都可以为自己的职员获取单独的邮件域名，并开发专门的网络应用软件，提供用户邮件传递服务，包括其特有的智能处理邮件的功能，例如安全服务功能等。必要时也可以为本单位用户计算机开发专用的用户代理软件，并同时将用户的其他邮箱邮件收发的功能组合进来。Foxmail 作为用户代理的成功在于用户可以在计算机上指定存放邮件存放的位置，这样即使系统重新安装，邮件数据存放区也不会被破坏。

不同于用户计算机浏览器 FTP 下载的文件复制，文件传输服务的网络应用市场潜力较大。这类应用市场主要是方便用户（用户很少干预并且操作方便）、性能优化（多线程下载以及负载转移）、断点续传（磁盘部分存储）以及可以增加特殊功能（针对企业数据传送

安全和其他功能)。单位使用时与电子邮件服务结合的趋势较大,尤其是开发符合单位安全机制和策略的文件传输和邮件传输服务的网络应用软件。

网络聊天是一个不断壮大的网络应用市场。这类软件充分发掘用户即时利用网络资源的潜能,将文字、图片、语音和视频聊天,以及文件传输、个人邮件、短信、个人娱乐、群体共享等综合到网络应用中。

最大的网络应用市场恐怕是网络游戏,这类游戏软件不断推陈出新。每款游戏软件结合了图像的动态艺术处理,游戏策划还包括参与游戏方的信息交流,网络编程只是其中较小的但非常重要的部分。目前,网络教育游戏的市场潜力更大,它的难点在于编程人员需要强化跨专业贯通和将知识游戏化处理的能力。

目前网络安全问题,尤其电子政务和电子商务安全问题,变得越来越突出。因此各种应用增加网络安全功能成为网络应用市场的一大亮点。通用的抗病毒软件,在病毒扫描时占用系统资源程度高,这样迫使应用软件开发时附带增加网络安全部分,避免过多的性能损失。

随着通信和计算机技术的提升以及网络应用性能要求的提高,网络基础设施和网络环境需要优化的网络软件支撑。网络运营商需要智能化的网络管理平台,提供网络管理、维护和参数配置,网络设备供应商需要开发专有路由选择算法和路由选择协议(例如,Cisco 公司路由器开发自己的路由选择协议),优化此类设备的网络环境。专用的网络需要开发不同安全特征的网关应用程序,提供机构内部的信息传输安全,IPv6 的引入需要开发更多网络应用程序、开发新的网络协议提高网络性能与安全。此外,还需要多计算机网络协同完成大型应用的功能实现。

随着信息时代信息化(提取信息)、数字化(传输)和网络化(分布与共享)需求的增长,网络应用程序的开发成为 IT 市场的主流。同时,随着网络应用的普及,网络应用逐步走向个性化、智能化和综合化。

1.2 Windows 环境下的应用程序

计算机上运行的网络应用程序首先是共享系统资源的应用程序,因此网络编程开发者需要了解 Windows 环境下的应用程序工作机制和编程开发环境问题。这一节将讨论在 Windows 环境下,应用程序的形态、运行环境以及相应的开发环境。

1.2.1 应用程序的形态

为什么双击扩展名为 exe 的文件之后它就会被 Windows 运行,而双击一个扩展名为 doc 的文件就被 Word 打开并显示其中的内容? 虽然 exe 文件的使用者是 Windows,而 doc 文件的使用者是 Word,但最终都是在 Windows 上运行应用程序。在操作系统中,应用程序可执行的代码最终被装入内存执行之前是以文件的方式存放在磁盘中的,也就是以可执行文件的方式。因此,应用程序在磁盘和在内存中是不同的。伴随着Windows操作系统的不断进步,可执行文件的格式也发生了巨大变化。最早在 DOS 中出现的是以 com 和以 exe 为扩展名的可执行文件,这一直是计算机黑客的最爱。以 com 为扩展名的可执行文件是 64kb 内存的 cp/m 机器的产物,最大 64K,内含 16 位程序的二进制代码映像,

没有重定位信息。在 VC++的 VC98\Include 目录里的 WINNT.h 文件中提供了 MZ (Mark Zbikowski,主要作者名字),NE(new executable),LE(linear executable)和 PE (portable executable)4 种 exe 可执行文件头部定义,如图 1-5 所示。它们都是在 MZ 格式基础上扩展起来的,具有内存重定位运行的能力,目前大多数应用程序的存储都是 PE 格式,详细说明见附录 1.2。NE 扩展用于 Win 3.x 的 exe 和 dll 分段可执行文件;LE 扩展用于 Win 3.x 和 Win 9x 所专有的 vxd 或者作为 OS/2 的 exe 的线性可执行文件格式;PE 扩展用于 Win 9x 和 Win NT/2000/XP 下的可移植的 32 位可执行文件。MZ 格式只能在 DOS 下运行,而它的扩展格式 NE,LE 和 PE 格式还可以运行于 Windows 界面,不过它们在不支持的 DOS 系统下运行时,利用 DOS Stub 程序运行显示错误信息。

图 1-5 是 exe 类可执行文件格式的物理分布,只有可执行文件装载到内存中才有可能被执行。PE 格式的可执行文件怎么变成运行中的应用进程? 也就是静态形式怎么变成动态形式的问题,这由 Windows 装载器解决。要保证可执行文件在内存中运行,需要解决 CPU 处理数据的字节顺序和内存地址映射两个问题。

图 1-5　exe 序列的可执行文件格式

采用术语"portable executable"是因为微软希望有一个能够在所有 Windows 平台上和所有 CPU 上通用的文件格式。从大的方面讲,这个目标已经实现。它适用于 NT 及其后代,95 及其后代和 CE。PE 文件中的段以某种顺序表示数据,而不同的计算机可能有不同的字节顺序格式。所谓的字节顺序指的是长度跨越多个字节的数据的存放形式,一般有 little endian 和 big endian 两种格式。例如,从内存地址 0x0000 开始有以下数据:

0x0000	0x12
0x0001	0x34
0x0002	0xab
0x0003	0xcd

对于读取一个地址为 0x0000 的 4 个字节变量,若字节顺序为 big-endian,则读出结果为 0x1234abcd;若字节顺序 little-endian,则读出结果为 0xcdab3412。如果 0x1234abcd 写入到以 0x0000 开始的内存中,则结果为

	big-endian	little-endian
0x0000	0x12	0xcd
0x0001	0x23	0xab
0x0002	0xab	0x34
0x0003	0xcd	0x12

现在主流的 CPU——Intel x86 系列是采用的 little endian 的格式存放数据,而 Motorola 系列的 CPU 采用的是 big endian。通过不同格式的 CPU 的处理,文件的数据在磁盘上的存储顺序是不同的,如 bmp,gif 文件采用 little endian 格式,而 jpeg 文件采用 big

endian 格式。这是文件数据装入内存必须考虑的问题。一种计算机的文件格式需要文件装载器进行特殊处理，才能在另一种格式的计算机上运行。

PE 格式的文件，作为一种顺序执行的文件装入到内存中时需要考虑文件地址、虚拟地址(VA)和相对虚拟地址(RVA)三种地址问题。文件地址，对应某个结构的地址(偏移)，比如用十六进制编辑器打开 PE 文件，看到的地址就是文件中的地址。虚拟地址就是程序中使用的地址，通常使用段地址——偏移量(offset)，不过 32 位系统使用单层(flat)内存模式，只需要考虑 32 位的偏移量。从逻辑上程序员被看成在一个 $2^{32}=4G$ 的段中写程序，实际上这个段构成的地址空间(64 位系统这个空间更大)是程序与操作系统共享的。程序可以申请这 2G～3G 的线性空间(NT 3.X 和 NT 4.0 为 2G 而 NT 4.0 企业版本为 3G)，而整个 4G 是由操作系统管理的(通过页表实现，分配线性地址空间就是分配页表给进程)。当 Windows 载入器将 PE 载入内存，在内存中它称作模块(Module)，文件从 hModule(模块句柄)这个地址(缺省值为文件映像基址 ImageBase)开始映射。记住这点：给你个 hModule，从那就知道一个数据结构(IMAGE_DOS_HEADER，见附录 1.2)，然后还可以知道所有的数据结构。相对虚拟地址是一个简单的相对于 PE 载入点(hModule)的内存偏移，PE 文件中许多字段都是用相对虚拟地址表示的，它为数据项的偏移量－从文件被映像进来的起点(基址)。例如，我们说 Windows 加载器将一个 PE 文件映像到虚拟地址空间的 0x400000 处，如果此 image 有一个表格开始于 0x401464，那么该表格的 RVA 就是 0x1464。PE 文件尽管有一个首选的载入地址 ImageBase，但是可以载入进程空间的任何地方，所以不能依赖于 PE 的载入点。由于这点，必须有一个方法来指定地址而不依赖于 PE 载入点的地址，这样避免把内存地址硬编码进 PE 文件，减少 PE 装载器的负担。因为每个模块都有可能被重载到任何虚拟地址空间，如果让 PE 装载器修正每个重定位项，这肯定是个梦魇。相反，如果所有重定位项都使用 RVA，那么 PE 装载器就不必操心那些东西了，即它只要将整个模块重定位到新的起始 VA。这就像相对路径和绝对路径的概念，RVA 类似相对路径，VA 就像绝对路径。注意，RVA 和 VA 是指内存中，不是指文件中。

下面将描述在内存中装载 PE 格式可执行文件的主要步骤：

(1) PE 文件被执行时，PE 装载器检查 DOS MZ header 里的 PE header 偏移量。对不支持的格式，运行 DOS Stub 程序提示用户。如果找到，则跳转到 PE header。

(2) PE 装载器检查 PE header 的有效性。如果有效，就跳转到 PE header 的尾部。

(3) 紧跟 PE header 的是段表。PE 装载器遍历整个段表，分配 4GB 的进程虚拟地址空间，将每个段加载到这个进程空间，读取其中的段信息，并采用文件映射方法将段映射到内存(磁盘与内存页交换，每页大小在 x86 CPU 上为 4k，需要的页放入内存，不再需要的回到磁盘)，同时附上段表里指定的属性。

(4) PE 装载器在内核中创建进程对象，如果有多线程则创建主线程对象以及其他内容。

(5) PE 装载器搜索 PE 文件中的引入表(import table)装载应用程序需要的动态链接库 DLL(dynamic link library)。对动态链接库的装载和对应用程序的装载方法类似。

(6) PE 装载器指向内存中 PE 文件首部所指定地址处的代码，应用程序主线程准备执行。