

可下载教学资料

<http://www.tup.tsinghua.edu.cn>



高等学校教材
计算机科学与技术

Linux 操作系统

原理与应用

陈莉君 康 华 编著



清华大学出版社

高等学校教材
计算机科学与技术

Linux 操作系统 原理与应用

陈莉君 康 华 编著

清华大学出版社
北京

内 容 简 介

本书从原理、设计思想的角度对 Linux 操作系统的核心内容进行全面的阐述,包括进程管理、中断、内核同步、内存管理、文件系统、I/O 设备管理等。书中各章还配以相应的实例分析及习题,便于读者把所学原理与实际开发结合起来,加深理解。在此基础上设立了专门的教学辅导网站,随时解答读者提出的问题。

本书对于希望深入 Linux 操作系统内部、阅读 Linux 内核源代码以及进行内核级程序开发的读者具有非常高的参考价值。本书可作为高等院校计算机相关专业的本科生、研究生的教材,同时, Linux 应用开发人员、嵌入式系统开发人员、系统程序员、系统管理员以及对 Linux 感兴趣的读者均可从本书的阅读中获益。

版权所有,翻印必究。举报电话:010-62782989 13501256678 13801310933

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

本书防伪标签采用特殊防伪技术,用户可通过在图案表面涂抹清水,图案消失,水干后图案复现;或将表面膜揭下,放在白纸上用彩笔涂抹,图案在白纸上再现的方法识别真伪。

图书在版编目(CIP)数据

Linux 操作系统原理与应用 / 陈莉君, 康华编著. —北京: 清华大学出版社, 2006. 1

(高等学校教材·计算机科学与技术)

ISBN 7-302-11834-5

I. L… II. ①陈… ②康… III. Linux 操作系统—高等学校—教材 IV. TP316.89

中国版本图书馆 CIP 数据核字 (2005) 第 106846 号

出版者: 清华大学出版社 地 址: 北京清华大学学研大厦
http://www.tup.com.cn 邮 编: 100084
社总机: 010-62770175 客户服务: 010-62776969

责任编辑: 付弘宇

印刷者: 北京市清华园胶印厂

装订者: 北京市密云县京文制本装订厂

发行者: 新华书店总店北京发行所

开 本: 185×260 印张: 16.5 字数: 392 千字

版 次: 2006 年 1 月第 1 版 2006 年 1 月第 1 次印刷

书 号: ISBN 7-302-11834-5/TP·7690

印 数: 1~4000

定 价: 24.00 元

编审委员会成员

(按地区排序)

清华大学	周立柱	教授
	覃 征	教授
	王建民	教授
	刘 强	副教授
	冯建华	副教授
北京大学	杨冬青	教授
	陈 钟	教授
	陈立军	副教授
北京航空航天大学	马殿富	教授
	吴超英	副教授
	姚淑珍	教授
中国人民大学	王 珊	教授
	孟小峰	教授
	陈 红	教授
北京师范大学	周明全	教授
北京交通大学	阮秋琦	教授
北京信息工程学院	孟庆昌	教授
北京科技大学	杨炳儒	教授
石油大学	陈 明	教授
天津大学	艾德才	教授
复旦大学	吴立德	教授
	吴百锋	教授
	杨卫东	副教授
华东理工大学	邵志清	教授
华东师范大学	杨宗源	教授
	应吉康	教授
东华大学	乐嘉锦	教授
上海第二工业大学	蒋川群	教授
浙江大学	吴朝晖	教授
	李善平	教授
南京大学	骆 斌	教授
南京航空航天大学	秦小麟	教授

南京理工大学	张功萱	教授
南京邮电学院	朱秀昌	教授
苏州大学	龚声蓉	教授
江苏大学	宋余庆	教授
武汉大学	何炎祥	教授
华中科技大学	刘乐善	教授
中南财经政法大学	刘腾红	教授
华中师范大学	王林平	副教授
	魏开平	教授
武汉理工大学	李中年	教授
国防科技大学	赵克佳	教授
	肖 依	副教授
中南大学	陈松乔	教授
湖南大学	林亚平	教授
	邹北骥	教授
西安交通大学	沈钧毅	教授
	齐 勇	教授
长安大学	巨永峰	教授
西安石油学院	方 明	教授
西安邮电学院	陈莉君	副教授
哈尔滨工业大学	郭茂祖	教授
吉林大学	徐一平	教授
	毕 强	教授
长春工程学院	沙胜贤	教授
山东大学	孟祥旭	教授
	郝兴伟	教授
山东科技大学	郑永果	教授
中山大学	潘小轰	教授
厦门大学	冯少荣	教授
福州大学	林世平	副教授
云南大学	刘惟一	教授
重庆邮电学院	王国胤	教授
西南交通大学	杨 燕	副教授

改革开放以来，特别是党的十五大以来，我国教育事业取得了举世瞩目的辉煌成就，高等教育实现了历史性的跨越，已由精英教育阶段进入国际公认的大众化教育阶段。在质量不断提高的基础上，高等教育规模取得如此快速的发展，创造了世界教育发展史上的奇迹。当前，教育工作既面临着千载难逢的良好机遇，同时也面临着前所未有的严峻挑战。社会不断增长的高等教育需求同教育供给特别是优质教育供给不足的矛盾，是现阶段教育发展面临的基本矛盾。

教育部一直十分重视高等教育质量工作。2001年8月，教育部下发了《关于加强高等学校本科教学工作，提高教学质量的若干意见》，提出了十二条加强本科教学工作提高教学质量的措施和意见。2003年6月和2004年2月，教育部分别下发了《关于启动高等学校教学质量与教学改革工程精品课程建设工作的通知》和《教育部实施精品课程建设提高高校教学质量和人才培养质量》文件，指出“高等学校教学质量和教学改革工程”是教育部正在制订的《2003—2007年教育振兴行动计划》的重要组成部分，精品课程建设是“质量工程”的重要内容之一。教育部计划用五年时间（2003—2007年）建设1500门国家级精品课程，利用现代化的教育信息技术手段将精品课程的相关内容上网并免费开放，以实现优质教学资源共享，提高高等学校教学质量和人才培养质量。

为了深入贯彻落实教育部《关于加强高等学校本科教学工作，提高教学质量的若干意见》精神，紧密配合教育部已经启动的“高等学校教学质量与教学改革工程精品课程建设工作”，在有关专家、教授的倡议和有关部门的大力支持下，我们组织并成立了“清华大学出版社教材编审委员会”（以下简称“编委会”），旨在配合教育部制定精品课程教材的出版规划，讨论并实施精品课程教材的编写与出版工作。“编委会”成员皆来自全国各类高等学校教学与科研第一线的骨干教师，其中许多教师为各校相关院、系主管教学的院长或系主任。

按照教育部的要求，“编委会”一致认为，精品课程的建设工作从开始就要坚持高标准、严要求，处于一个比较高的起点上；精品课程教材应该能够反映各高校教学改革与课程建设的需要，要有特色风格、有创新性（新体系、新内容、新手段、新思路，教材的内容体系有较高的科学创新、技术创新和理念创新的含量）、先进性（对原有的学科体系有实质性的改革和发展、顺应并符合新世纪教学发展的规律、代

表并引领课程发展的趋势和方向)、示范性(教材所体现的课程体系具有较广泛的辐射性和示范性)和一定的前瞻性。教材由个人申报或各校推荐(通过所在高校的“编委会”成员推荐),经“编委会”认真评审,最后由清华大学出版社审定出版。

目前,针对计算机类和电子信息类相关专业成立了两个“编委会”,即“清华大学出版社计算机教材编审委员会”和“清华大学出版社电子信息教材编审委员会”。首批推出的特色精品教材包括:

(1) 高等学校教材·计算机应用——高等学校各类专业,特别是非计算机专业的计算机应用类教材。

(2) 高等学校教材·计算机科学与技术——高等学校计算机相关专业的教材。

(3) 高等学校教材·电子信息——高等学校电子信息相关专业的教材。

(4) 高等学校教材·软件工程——高等学校软件工程相关专业的教材。

(5) 高等学校教材·信息管理与信息系统。

清华大学出版社经过近 20 年的努力,在教材尤其是计算机和电子信息类专业教材出版方面树立了权威品牌,为我国的高等教育事业做出了重要贡献。清华版教材经过 20 多年的精雕细刻,形成了技术准确、内容严谨的独特风格,这种风格将延续并反映在特色精品教材的建设中。

清华大学出版社教材编审委员会
E-mail: dingl@tup.tsinghua.edu.cn

芬兰大学生 Linus Torvalds 在赫尔辛基大学学习操作系统课程时，由于不满足于使用教学用操作系统 Minix，他从着手开发一个简单的程序开始，逐步开发了显示器、键盘和调制解调器的驱动程序，最后编写了磁盘驱动程序、文件系统，一个操作系统的原型就这样形成了。

这个诞生于学生之手的 Linux，在 Internet 这片肥沃的土壤中不断成长，逐步发展为与 Unix、Windows 并驾齐驱的实用操作系统。Linux 与 Windows 不同而与 Unix 外表相似，它的窗口向所有人完全敞开，任何想了解其内在机理的爱好者都可以走进其内部世界。

1999 年春，我们有幸走进了这个开放的世界，那时分析的是 Linux 内核 2.0 版。在阅读源代码的基础上，我们编写了《Linux 操作系统内核分析》一书。随着 Linux 内核版本的不断更新，我们又陆续编写和翻译了针对 Linux 内核 2.2、2.4 及 2.6 版本的相关书籍。

Linux 内核是由 C 语言和汇编语言编写的（以 C 语言为主），其全部源代码是一个庞大的集合。如何在这个庞大而复杂的集合中抓住主要内容，找到进入 Linux 内部的突破口，如何使 Linux 的源代码适应自己的需求，并在此基础上进行内核级程序的开发，这是本书要探讨的内容。

首先，本书的第 1 章引导读者初识操作系统，从操作系统这座大厦的各个侧面观其外貌，并走近 Linux 内核，考察其内在结构。但这种全局性的认识仅是个起点，要从根本上了解操作系统，与之相关的硬件知识是不可或缺的。第 2 章以 Intel x86 处理器为基础，从内存寻址的角度介绍了硬件对操作系统特有的支持，这种支持使得虚拟内存管理有了坚实的物质基础。

进程管理是操作系统的灵魂。第 3 章从内核实现的角度分析了进程赖以存活的各种数据结构，但没有过多涉及具体代码。进程是一个动态变化的实体，这一章从生命历程的角度说明了进程从诞生到死亡的过程。

内存作为计算机系统的重要资源，因为其容量的有限和程序规模的不断扩大，需要从技术上对其容量进行扩充。第 4 章主要讨论了虚拟内存管理的实现技术，并给出了一个实例，帮助读者加深理解。

随后的几章（5~9 章）对中断、内核同步、系统调用、文件系统以及驱动程序

从内核设计的角度进行了讨论，并给出相应的实例，加深读者对相关内容的理解。

本书的附录部分简单介绍了内核链表和内核模块的概念。最后给出了一些网络资源的地址，感兴趣的读者可以在相关网站上获取更丰富的资料。

为了突出主题，本书尽量简化相关内容。但为了填补课堂教学和实践开发之间的鸿沟，我们将在 www.kerneltravel.net 网站上辅助本书的出版，不定期发布专题性电子刊物。网站上也将针对学生学习操作系统课程后苦于无用武之地的现状，讨论如何进行 Linux 内核层面上的系统软件开发。并配以有实用价值或指导意义的实验，帮助学生认识 Linux 内核，学习 Linux 内核，进而理解与开发 Linux 内核。同时，网站提供本书和实验中涉及的源代码，并将建立一个以教学为中心的论坛，以解答读者的疑问。

编 者

2005 年 4 月

第 1 章 操作系统概述	1
1.1 认识操作系统	1
1.1.1 从使用者角度	1
1.1.2 从程序开发者角度	2
1.1.3 从操作系统在整个计算机系统中所处位置	2
1.1.4 从操作系统设计者的角度	3
1.2 操作系统的发展	4
1.2.1 操作系统的演变	4
1.2.2 硬件的发展轨迹	5
1.2.3 软件的轨迹	6
1.2.4 单内核与微内核操作系统	7
1.3 开放源代码的 Unix/Linux 操作系统	8
1.3.1 Unix 的诞生和发展	8
1.3.2 Linux 的诞生	9
1.3.3 操作系统标准 POSIX	9
1.3.4 GNU 和 Linux	9
1.3.5 Linux 的开发模式	10
1.4 Linux 内核	10
1.4.1 Linux 内核的位置	10
1.4.2 Linux 内核的作用	11
1.4.3 Linux 内核子系统	11
1.5 Linux 内核源代码	13
1.5.1 多版本的内核源代码	13
1.5.2 Linux 内核源代码的结构	13
1.5.3 Linux 内核源代码分析工具	14
习题 1	15
第 2 章 内存寻址	17
2.1 内存寻址简介	17
2.1.1 Intel x86 CPU 寻址方式的演变	18
2.1.2 IA32 寄存器简介	19

2.1.3	物理地址、虚拟地址及线性地址	21
2.2	分段机制	22
2.2.1	地址转换及保护	24
2.2.2	Linux 中的段	24
2.3	分页机制	25
2.3.1	页与页表	25
2.3.2	线性地址到物理地址的转换	28
2.3.3	分页示例	28
2.3.4	页面高速缓存 (cache)	29
2.3.5	Linux 中的分页机制	30
2.4	Linux 中的汇编语言	31
2.4.1	AT&T 与 Intel 汇编语言的比较	31
2.4.2	AT&T 汇编语言的相关知识	32
2.5	Linux 系统地址映射示例	33
	习题 2	35
第 3 章	进程	37
3.1	进程介绍	37
3.1.1	程序和进程	37
3.1.2	进程的层次结构	38
3.1.3	进程状态	39
3.1.4	进程实例	40
3.2	进程控制块	41
3.2.1	进程状态	42
3.2.2	进程标识符	43
3.2.3	进程之间的亲属关系	43
3.2.4	进程控制块的存放	44
3.3	进程的组织方式	45
3.3.1	进程链表	45
3.3.2	散列表	46
3.3.3	可运行队列	47
3.3.4	等待队列	47
3.4	进程调度	48
3.4.1	基本原理	48
3.4.2	时间片	50
3.4.3	Linux 进程调度时机	50
3.4.4	进程调度的依据	51
3.4.5	调度函数 <code>schedule()</code> 的实现	52

3.5	进程的创建	54
3.5.1	创建进程	55
3.5.2	线程及其创建	56
3.6	与进程相关的系统调用及其应用	58
3.6.1	fork 系统调用	58
3.6.2	exec 系统调用	59
3.6.3	wait 系统调用	60
3.6.4	exit 系统调用	62
3.6.5	进程的一生	63
3.7	与调度相关的系统调用及应用	63
	习题 3	65
第 4 章	内存管理	67
4.1	Linux 的内存管理概述	67
4.1.1	虚拟内存、内核空间和用户空间	67
4.1.2	虚拟内存实现机制间的关系	69
4.2	进程用户空间的管理	70
4.2.1	进程用户空间的描述	71
4.2.2	进程用户空间的创建	74
4.2.3	虚存映射	76
4.2.4	进程的虚存区示例	76
4.2.5	与用户空间相关的系统调用	78
4.3	请页机制	79
4.3.1	缺页异常处理程序	79
4.3.2	请求调页	81
4.3.3	写时复制	83
4.4	物理内存的分配与回收	83
4.4.1	伙伴算法	85
4.4.2	物理页面的分配	86
4.4.3	物理页面的回收	88
4.4.4	slab 分配模式	89
4.4.5	内核空间非连续内存区的分配	93
4.5	交换机制	95
4.5.1	交换的基本原理	95
4.5.2	页面交换守护进程 kswapd	99
4.6	内存管理实例	99
4.6.1	相关背景知识	100
4.6.2	代码体系结构介绍	100

4.6.3	实现步骤	103
4.6.4	程序代码	103
习题 4		108
第 5 章	中断和异常	110
5.1	中断的基本知识	110
5.1.1	中断向量	110
5.1.2	外设可屏蔽中断	111
5.1.3	异常及非屏蔽中断	112
5.1.4	中断描述符表	112
5.1.5	相关汇编指令	113
5.2	中断描述符表的初始化	114
5.2.1	IDT 表项的设置	114
5.2.2	对陷阱门和系统门的初始化	115
5.2.3	中断门的设置	116
5.3	中断处理	116
5.3.1	中断和异常的硬件处理	116
5.3.2	中断请求队列的建立	117
5.3.3	中断处理程序的执行	119
5.3.4	从中断返回	121
5.4	中断的下半部处理机制	121
5.4.1	为什么把中断分为两部分来处理	122
5.4.2	小任务机制	122
5.4.3	下半部	124
5.4.4	任务队列	125
5.5	中断应用——时钟中断	125
5.5.1	时钟	125
5.5.2	时钟运作机制	126
5.5.3	Linux 的时间系统	127
5.5.4	时钟中断处理程序	128
5.5.5	时钟中断的下半部处理	129
5.5.6	定时器及其应用	129
习题 5		132
第 6 章	系统调用	133
6.1	系统调用与应用编程接口、系统命令、内核函数的关系	133
6.1.1	系统调用与 API	133
6.1.2	系统调用与系统命令	134

6.1.3 系统调用与内核函数	134
6.2 系统调用处理程序及服务例程	135
6.2.1 初始化系统调用	136
6.2.2 system_call()函数	136
6.2.3 参数传递	137
6.2.4 跟踪系统调用的执行	139
6.3 封装例程	140
6.4 添加新系统调用	141
6.5 实例——利用系统调用实现一个调用日志收集系统	143
6.5.1 代码体系结构	143
6.5.2 把代码集成到内核中	146
6.5.3 实现步骤	148
习题 6	148
第 7 章 内核中的同步	149
7.1 临界区和竞争状态	149
7.1.1 临界区举例	149
7.1.2 共享队列和加锁	150
7.1.3 确定保护对象	151
7.1.4 死锁	152
7.1.5 并发执行的原因	153
7.2 内核同步方法	153
7.2.1 原子操作	153
7.2.2 自旋锁	155
7.2.3 信号量	156
7.3 并发控制实例	157
7.3.1 内核任务及其并发关系	158
7.3.2 实现机制	158
7.3.3 关键代码解释	162
7.3.4 实现步骤	163
习题 7	164
第 8 章 文件系统	165
8.1 Linux 文件系统基础	165
8.1.1 Linux 文件结构	165
8.1.2 Linux 文件系统	166
8.1.3 文件类型	167
8.1.4 文件访问权限	168

8.2	虚拟文件系统	168
8.2.1	虚拟文件系统的引入	168
8.2.2	VFS 中的数据结构	170
8.2.3	VFS 超级块数据结构	171
8.2.4	VFS 的索引节点	173
8.2.5	目录项对象	174
8.2.6	与进程相关的文件结构	176
8.2.7	主要的数据结构之间的关系	179
8.3	文件系统的注册、安装与卸载	180
8.3.1	文件系统的注册和注销	180
8.3.2	文件系统的安装	181
8.3.3	文件系统的卸载	183
8.4	页缓冲区	183
8.4.1	address_space 对象	183
8.4.2	address_space 对象的操作函数表	184
8.5	文件的打开与读写	185
8.5.1	打开文件	185
8.5.2	读写文件	187
8.6	编写一个文件系统	189
8.6.1	Linux 文件系统的实现要素	189
8.6.2	什么是 romfs 文件系统	191
8.6.3	romfs 文件系统的布局与文件结构	191
8.6.4	具体实现的对象	192
	习题 8	195
第 9 章	设备驱动	196
9.1	概述	196
9.2	设备驱动程序基础	198
9.2.1	I/O 端口	199
9.2.2	设备文件	200
9.2.3	中断处理	201
9.2.4	设备驱动程序框架	203
9.3	字符设备驱动程序	204
9.3.1	字符设备驱动程序的注册	204
9.3.2	简单的字符设备驱动程序示例	205
9.4	块设备驱动程序	208
9.4.1	块设备驱动程序的注册	209
9.4.2	块设备请求	212

习题 9	215
附录 A 内核中的链表	216
A.1 链表数据结构简介	216
A.2 内核链表数据结构的定义及初始化	217
A.3 操作链表的接口	218
A.4 遍历链表	219
附录 B 内核模块	221
B.1 什么是模块	221
B.2 编写一个简单的模块	221
B.3 模块编程的基础知识	222
B.4 模块的编译	224
B.5 模块实用程序 modutils	226
附录 C Linux 内核编译	228
C.1 内核简介	228
C.2 为什么重新编译内核	228
C.3 内核编译模式	229
C.4 新版本内核的获取和更新	229
C.5 内核编译	230
C.6 修改并重启管理器	232
附录 D Linux 编程基础 (C 语言环境)	233
D.1 Linux 编程常识	233
D.1.1 相关标准 (ANSI C、POSIX、SVID、XPG)	233
D.1.2 函数库和系统调用	234
D.1.3 在线文档 (man、info、HOWTO)	235
D.1.4 C 语言编程风格	237
D.2 Linux 上的 C/C++ 编译器和调试器	238
D.2.1 运行 gcc/egcs	238
D.2.2 gcc/egcs 的主要选项	240
D.2.3 gdb 简介	240
D.2.4 gdb 的常用命令	241
D.2.5 gdb 使用示例	242
D.3 GNU make 和 makefile	243
D.3.1 GNU make	243
D.3.2 makefile 的基本结构	243

D.3.3	makefile 的变量	244
D.3.4	GNU make 的主要预定义变量	245
D.3.5	GNU make 的隐含规则	245
D.3.6	运行 make	246
参考文献		247