

自由软件丛书

Linux

核心源代码 分析

- 彻底研究自由软件
- 探索Linux核心的秘密

彭晓明 王强
编著

人民邮电出版社
www.pptph.com.cn

自由软件丛书

Linux 核心源代码分析

彭晓明 王 强 编著

人民邮电出版社

内 容 提 要

本书以最新版 Linux 核心为基础,通过对源代码的分析,全面深入地介绍 Linux 核心的工作原理和实现技术。主要内容包括:进程管理、内存管理、文件系统、网络、P 设备管理、核心机制、核心模组、开关及过程、核心编译等,另外对 PCI 设备工作原理、中断管理以及 Linux 核心所特有的一些技术也作了较为深入的介绍。

本书适合于从事 Linux 研究与开发的人员阅读。对于大专院校的高年级本科生和研究生来说,本书也是一本较为理想的教材。

JS.../13

自由软件丛书

Linux 核心源代码分析

- ◆ 编 著 彭晓明 王 强
责任编辑 张瑞喜
- ◆ 人民邮电出版社出版发行 北京市崇文区夕照寺街 14 号
邮编 100061 电子函件 315@pptph.com.cn
网址 <http://www.pptph.com.cn>
北京鸿佳印刷厂印刷
新华书店总店北京发行所经销
- ◆ 开本:787×1092 1/16
印张:15.5
字数:378 千字
印数:1-6 000 册

2000 年 8 月第 1 版

2000 年 8 月北京第 1 次印刷

ISBN 7-115-08662-1/TP·1736

定价:24.00 元

丛书前言

自由软件的出现，改变了传统的以公司为主体的封闭式的软件开发模式。采用了开放和协作的开发模式，无偿提供源代码，允许任何人取得、修改和重新发布自由软件的源代码。这种开发模式激发了世界各地的软件开发人员的积极性和创造热情。大量软件开发人员投入到自由软件的开发中。软件开发人员的集体智慧得到充分发挥，大大减少了不必要的重复劳动，并使自由软件的脆弱点能够及时被发现和克服。任何一家公司都不可能投入如此强大的人力去开发和检验商品化软件。这种开发模式使自由软件具有强大的生命力。

就目前我国计算机软件状况而言，系统软件和大部分应用软件平台基本上被国外软件公司所垄断，民族软件产业的发展面临着极大的困难。自由软件无保留地提供源代码，使我们可以在高起点起步，非常有利于打破垄断；有利于我国软件行业在较短的时间内彻底改变目前被动的局面。自由软件的免费和自由传播特性十分适合于我国目前的经济状况。

适应这样一种形势，我们组织编写了这套自由软件丛书。丛书以介绍最新自由软件的技术和使用技巧为主。自由软件的缺点是缺乏开发公司的技术支持，文档通常也不齐全，给软件的使用带来了较大的不便，丛书正好弥补这方面的不足。丛书不仅仅涉及 Linux 操作系统，而且还涉及与之相关的网络服务器、数据库、多媒体等众多自由软件以及自由软件开发技术。力图较为全面地介绍和讲解自由软件的相关技术。

希望这套丛书的出版，能够推进自由软件在我国的发展进程，能够给广大的软件工作者的学习和研究带来一定的帮助。同时热切期待广大读者对丛书提出宝贵意见，也欢迎读者参与丛书的编写，让我们共同努力，为自由软件在我国生根、开花、结果做出贡献。

“自由软件丛书”编委会

前 言

以 Linux 为杰出代表的自由软件，遵循源代码开放、自由使用、自由传播的宗旨，为我国软件工作者学习世界最先进的软件技术提供了良好的条件，也为我国民族软件产业的发展提供了良好的契机。随着 Linux 操作系统在我国的普及，越来越多的软件工作者希望能够深入地了解 Linux 核心的工作原理和实现算法，以便使 Linux 更符合我国国情，继而开发出具有自主知识产权的操作系统软件。

本书试图通过对 Linux 核心源代码的分析，将 Linux 核心的工作原理和实现技术展现给读者。本书的内容基本上涵盖了 Linux 核心的各个方面，不仅介绍基本工作原理，还分析了具体的实现技术和代码。

本书是作者在近年来分析和学习 Linux 的基础上参考众多的文献资料编写而成的。主要以 x86 微机系统为背景，以 Linux 核心源代码的数据结构为线索，分析其工作原理和实现技术。第一章主要对 Linux 核心的技术特点进行概括性的描述，以帮助读者建立对 Linux 核心技术的整体认识。第二章主要介绍进程管理技术，内容包括 Linux 进程管理技术的特点、主要数据结构、进程调度、进程控制、进程间通信技术，还对进程与文件和虚拟内存的关系作了简要的说明。第三章主要介绍内存管理技术，先介绍虚拟内存模型和 80386 内存管理机构，然后着重分析请求页面虚拟内存管理技术的实现细节。第四章主要介绍文件系统，先介绍 Linux 文件系统的历史和技术特点，然后重点分析虚拟文件系统和 EXT2 文件系统，最后对缓存、内存守护进程等方面的技术作简要分析。第五章以 TCP/IP 为重点分析 Linux 的网络系统，主要内容包括 BSD、INET 套接字、IP 层、地址解析协议和 IP 路由。第六章介绍 PCI 总线的结构、PCI 桥和 PCI 初始化技术。第七章分析设备管理技术，着重讨论设备驱动程序、硬盘、网络设备以及中断管理等技术。第八章介绍核心机制。第九章介绍核心模组。第十章介绍 Linux 开机过程，从另外一个方面帮助读者加深对核心的理解。第十一章介绍核心的配置和编译技术。

彭晓明、孟加负责编写本书第一、二、四、五、七、十、十一章，王强负责编写第三、六、八、九章，彭晓明负责全书统稿，李红侠、赵善松、李玉珍、郑玉强、黄卿贤、康蓓蕾等也为本书的编写做了大量的工作。

最后对所有关心、支持和帮助过本书编写的领导、朋友和家人表示诚挚的谢意！

由于作者的能力和水平有限，加上时间仓促，本书可能会有不妥和错误之处，如能给予指正将不胜感谢！

作者的电子信箱地址为：pengxm@263.net

作者

2000 年 4 月

目 录

第一章 概论	1
1.1 Linux 简介	2
1.1.1 Linux 的诞生	2
1.1.2 自由软件运动与 Linux	3
1.1.3 Linux 在中国	4
1.2 Linux 的技术特点	5
1.2.1 Linux 的特点	5
1.2.2 Linux 的组成	6
1.2.3 Linux 的不足	8
1.3 核心源代码	9
1.3.1 源代码的获取	9
1.3.2 源代码的结构	10
1.3.3 源代码的阅读	11
第二章 进程管理	15
2.1 进程概述	16
2.1.1 进程的定义	16
2.1.2 进程的状态	17
2.1.3 进程控制块	18
2.1.4 进程控制	18
2.1.5 进程间通信	19
2.1.6 进程调度	19
2.2 Linux 的进程	20
2.2.1 主要数据结构	20
2.2.2 进程表的主要域	23
2.2.3 进程执行模式	25
2.2.4 进程标识符	25
2.3 进程调度	27
2.3.1 调度参数	27
2.3.2 调度工作	28



- 2.3.3 多处理机进程调度 29
- 2.3.4 调度的实现 29
- 2.4 进程的创建和撤销 31
 - 2.4.1 进程的创建 31
 - 2.4.2 进程的撤销 31
 - 2.4.3 程序的执行 32
- 2.5 文件 35
- 2.6 虚拟内存 38
- 2.7 计时和定时器 41
- 2.8 信号 43
 - 2.8.1 信号的类型 43
 - 2.8.2 信号的处理 43
- 2.9 管道 45
 - 2.9.1 管道的实现 45
 - 2.9.2 管道的操作 46
 - 2.9.3 命名管道 46
- 2.10 System V 进程间通信机制 47
 - 2.10.1 主要数据结构 47
 - 2.10.2 消息队列 47
 - 2.10.3 信号灯 49
 - 2.10.4 共享内存 52

第三章 内存管理 55

- 3.1 内存管理技术概述 56
- 3.2 虚拟内存模型 59
 - 3.2.1 页面请求 60
 - 3.2.2 页面交换 61
 - 3.2.3 共享内存 61
 - 3.2.4 物理和虚拟地址模式 62
 - 3.2.5 访问控制 62
 - 3.2.6 缓存 62
- 3.3 80386 内存管理机构 64
 - 3.3.1 内存组织 64
 - 3.3.2 虚拟地址的转换 65
 - 3.3.3 线性地址的转换 66
 - 3.3.4 翻译后援存储器 67
 - 3.3.5 内存保护 68

3.4 主要数据结构 69

3.5 页面管理 71

 3.5.1 页面分配 71

 3.5.2 页面的回收 71

 3.5.3 内存映射 72

3.6 页面装卸 74

 3.6.1 页面请求 74

 3.6.2 页面缓存 74

 3.6.3 页面换出和淘汰 75

 3.6.4 交换缓存 78

 3.6.5 页面装入 78

第四章 文件系统 81

4.1 文件系统简介 82

 4.1.1 文件 82

 4.1.2 文件目录 84

 4.1.3 路径名 85

4.2 Linux 文件系统概述 86

 4.2.1 历史 86

 4.2.2 几个概念 87

4.3 EXT2 文件系统 89

 4.3.1 物理布局 89

 4.3.2 i-节点 90

 4.3.3 超级块 91

 4.3.4 块组描述子 93

 4.3.5 目录 93

 4.3.6 文件检索 94

 4.3.7 数据块的管理 94

4.4 虚拟文件系统 96

 4.4.1 VFS 概述 96

 4.4.2 VFS 超级块 97

 4.4.3 VFS i-节点 99

 4.4.4 注册文件系统 101

 4.4.5 装卸文件系统 102

 4.4.6 VFS i-节点缓存 104

 4.4.7 目录缓存 105

4.5 缓冲区缓存及其它 106



4.5.1 缓冲区缓存 106
4.5.2 bdflush 核心守护进程 107
4.5.3 update 进程 108
4.5.4 /proc 文件系统 108
4.5.5 特殊设备文件 108

第五章 网络 109

5.1 TCP/IP 网络基础 110
5.1.1 协议簇结构 110
5.1.2 IP 地址 111
5.1.3 地址类别 111
5.1.4 子网 112
5.1.5 无类地址和 CIDR 113
5.1.6 主机名 114
5.1.7 数据包传输 114
5.1.8 Linux 网络层 116
5.2 BSD 套接字接口 117
5.3 INET 套接字层 119
5.3.1 创建 BSD 套接字 119
5.3.2 绑定地址 121
5.3.3 建立连接 122
5.3.4 监听 122
5.3.5 接收连接请求 123
5.4 IP 层 124
5.4.1 套接字缓冲区 124
5.4.2 接收 IP 数据包 128
5.4.3 发送 IP 数据包 128
5.4.4 数据包分段 129
5.5 其它 130
5.5.1 地址解析协议 130
5.5.2 IP 路由 130

第六章 PCI 总线 133

6.1 PCI 总线及其结构 134



6.1.1 PCI 总线的结构	134
6.1.2 PCI 地址空间	135
6.1.3 PCI 配置信息	135
6.1.4 PCI I/O 端口及存储器地址	136
6.2 PCI 桥	138
6.2.1 PCI-ISA 桥	138
6.2.2 PCI-PCI 桥	138
6.3 PCI 初始化	141
6.3.1 数据结构	141
6.3.2 PCI 驱动程序	143
6.3.3 PCI BIOS 函数	146
6.3.4 PCI 修正	146
第七章 设备管理	149
7.1 设备概述	150
7.1.1 设备种类	150
7.1.2 设备控制器	150
7.2 设备驱动程序	151
7.2.1 查询和中断	152
7.2.2 直接内存访问	152
7.2.3 内存	153
7.2.4 核心接口	153
7.3 硬盘	157
7.3.1 IDE 硬盘	158
7.3.2 IDE 子系统的初始化	159
7.3.3 SCSI 硬盘	160
7.3.4 SCSI 子系统的初始化	161
7.3.5 块设备请求	162
7.4 网络设备	164
7.4.1 device 数据结构	164
7.4.2 网络设备的初始化	165
7.5 中断	166
7.5.1 关于中断	166
7.5.2 可编程中断控制器	166
7.5.3 中断初始化	167
7.5.4 中断处理	168



第八章 核心机制 169

- 8.1 底半处理 170
 - 8.1.1 数据结构 170
 - 8.1.2 处理过程 171
- 8.2 任务队列 172
- 8.3 定时器 174
- 8.4 等待队列 176
- 8.5 Buzz 锁 177
- 8.6 信号量 178

第九章 核心模组 181

- 9.1 模组简介 182
- 9.2 模组的装入 184
 - 9.2.1 装入方式 184
 - 9.2.2 数据结构 184
 - 9.2.3 insmod 命令 186
- 9.3 模组的卸出 188

第十章 开机、关机 189

- 10.1 开机过程 190
- 10.2 核心启动 192
- 10.3 inittab 196
- 10.4 rc.sysinit 199
- 10.5 rcN.d 209
- 10.6 关机 215

第十一章 核心编译 217

- 11.1 关于核心编译 218
 - 11.1.1 编译目的 218
 - 11.1.2 准备工作 218

11.1.3 编译步骤	220
11.2 配置核心	222
11.2.1 配置内容	222
11.2.2 配置过程	223
11.3 核心编译和启用	233
11.3.1 编译	233
11.3.2 启用	233
11.3.3 利用补丁升级	234

第一章

概论

Linux 是一种自由的 UNIX 类多用户、多任务操作系统，可运行在 Inter 80386 及更高档次的 PC 机、ARMs、DEC Alpha、SUN Sparc、M68000、MIPS 和 PowerPC 等多种计算机平台，已成为应用广泛、可靠性高、功能强大的计算机操作系统。

本章的主要内容包括：

- ◆Linux 简介
- ◆Linux 技术特点
- ◆Linux 核心源代码



1.1 Linux 简介

Linux 是一种自由的 UNIX 类多用户、多任务操作系统，可运行在 Inter 80386 及更高档次的 PC 机、ARMs、DEC Alpha、SUN Sparc、M68000、MIPS 和 PowerPC 等多种计算机平台，已成为应用广泛、可靠性高、功能强大的计算机操作系统。

1.1.1 Linux 的诞生

谈到 Linux，得从 UNIX 说起。UNIX 是 Thompson 和 Ritchie 于 1974 年在《ACM 通讯》中的一篇文章中首次提出的。自问世以来，UNIX 迅速在世界范围内推广，不仅成为高档微机、工作站、小型机的主流操作系统，而且已进入中、大型计算机领域，成为“事实”上的标准操作系统。早期的 UNIX 操作系统基本上都是商品化的软件产品，其价格相当昂贵，这在一定程度上影响了 UNIX 的普及。为了使更多的用户能够得益于 UNIX 强大的功能，许多可以自由使用、自由传播的免费 UNIX 应运而生。其中之一是 Minix。虽然 Minix 不完全具备现代 UNIX 的许多特性，但由于它是一种小巧的操作系统，且可以用于 PC 机，因此受到了人们的广泛关注。许多软件工作者开始为 Minix 增加新的功能，开发新的应用程序，或以 Minix 为基础，开发新的自由 UNIX。

在研究和改进 Minix 过程中，芬兰青年 Linus 草创了 Linux 操作系统，从而开创了 Linux 的历史。Linus 在自己的 PC 机上，利用 Minix 为开发平台，开发了属于他自己的第一个程序。这个程序包括两个进程，向屏幕上写字母，然后用一个定时器来切换这两个进程。刚开始的时候，Linus 根本没有想到要编写一个操作系统核心。1991 年，他需要一个简单的终端仿真程序来存取 Usenet 新闻组的内容，于是他就在前两个草草编写的进程的基础上又写了一个程序。同年，Linus 觉得还需要从网上下载某些文件，于是他又编写了一个磁盘驱动程序和一个文件系统。这样，Linux 就以一种极其不同寻常但也极其自然的方式问世了。

为了不至于使这个羽毛未丰的操作系统马上夭折，Linus 并没有在 Minix 新闻组中公布它，而是在赫尔辛基技术大学的一台 FTP 服务器上发了一则消息，说用户可以下载 Linux 的公开版本。到 1992 年 1 月止，全世界大约有 100 人在使用 Linux。虽然这时认识 Linux 的人比较少，但正是他们为 Linux 作了关键性的在线洗礼。他们所提供的所有初期上载代码和评论后来被证明对 Linux 的发展起到了至关重要的作用。从此，网上的任何人在任何地方都可以得到 Linux 的基本文件，并可通过电子邮件发表评论或者提供修正代码，Usenet 还专门为它开辟了一个论坛。于是，Linux 就从最初的个人思想产品变成了由无数 Linux 爱好者发起的一场运动。

实际上，Linux 的成熟与发展是 Internet 的产物，也是 Internet 时代创造的一个奇迹。

1.1.2 自由软件运动与 Linux

1. 自由软件运动

为了摆脱商业软件公司对软件特别是对操作系统软件的控制，自 1984 年起，麻省理工学院（MIT）开始支持 Richard Stallman 的努力，即在软件开发团体中发起自由软件运动，从而导致了自由软件基金会（FSF, Free Software Foundation）的建立及 GPL 协议和 GNU 项目的诞生。Richard Stallman 在其它许多人的协作下，制定了通用公共许可协议（GPL, General Public License）。与传统的商业软件许可协议不同，GPL 保证任何人有共享和修改自由软件的自由。任何人有权取得、修改和重新发布自由软件的源代码，并且规定在不增加附加费用的条件下可以得到自由软件的源代码（基本的发布费用除外）。这一规定保证了自由软件总的费用是低的，在使用 Internet 的情况下，则是完全免费的。GPL 对推动自由软件发展起了重要的作用。

GPL 协议还规定自由软件的衍生作品必须以 GPL 作为它重新发布的许可协议。这一规定保证了自由软件及其衍生作品继续保持自由状态。

GPL 协议也不反对销售自由软件，为商业公司介入自由软件事业敞开了大门。商业公司的介入弥补了自由软件的不足，对推动自由软件的普及起了很大的作用。

自由软件基金会发起人的主要项目是 GNU，其目标是建立可自由发布和可移植的 UNIX 类操作系统。在开始实施 GNU 项目的时候，没有多少高质量的自由软件可供项目使用，所以，愿意为 GNU 项目做出贡献的人们先从系统的应用软件和工具入手。因为 GPL 也是自由软件基金会发布的，所以，GNU 操作系统的许多关键组成部分都置于 GPL 协议的约束之下。GNU 项目本身产生的主要软件包括：Emacs 编辑软件、gcc 编译软件、bash 命令解释程序和编程语言，以及 gawk（GNU's awk）等，此外还有其它许多操作系统必不可少的工具。

除了按 GPL 发布的自由软件之外，还有许多按其它许可协议发布的自由软件，如 X Windows 系统、TEX 排版系统和 Perl 语言等。随着时间的推移，GNU 项目将这些软件也包括进来。

2. GPL Linux

Linux 的第一个“产品”版 Linux 1.0 于 1993 年问世，按完全自由扩散版权进行扩散。Linus 要求所有的源码必须公开，而且任何人均不得从 Linux 交易中获利，这说明 Linus 是一个典型的理想主义者。然而半年以后，他开始意识到，这种纯粹的自由软件的理想对于 Linux 的普及和发展实际上是一种障碍，因为它限制了 Linux 以磁盘或 CD-ROM 等媒体形式进行扩散的可能，也限制了一些商业公司参与 Linux 的进一步开发并提供技术支持的良好愿望，于是 Linus 决定将 Linux 转向 GPL。

这一转变后来被证明对于 Linux 的进一步发展是极为重要的。从此以后，多家技术力量雄厚又善于市场运作的商业软件公司加入到原先完全由业余爱好者和网络黑客所参与的这场自由软件运动，开发出了多种 Linux 的发行版本（distribution），并增加了更易于用户使用的图形界面和众多的软件开发工具，极大地拓展了 Linux 的用户基础。Linus 本人

也认为：“使 Linux 成为 GPL 的一员是我一生中做过最漂亮的一件事。”

目前用户使用的 Linux 大部分是由商业软件公司提供的，如著名的 RedHat Linux、Slackware Linux 和 Caldera Linux 等。我国的一些软件公司和研究机构，如实达朗新公司，拓林思（中国）公司、北京冲浪平台公司、中科院软件研究所等，也加入到 Linux 发行者的行列之中，为我国用户提供了中文化的 Linux，如 Tom Linux、Turbo Linux、Xteam Linux、红旗 Linux 等。这些发行版本的价格非常低廉，基本上只包含发行费和介质费，而内容却极为丰富。以 Tom Linux 为例，其价格只有区区的几十元人民币，但却包含了：Linux 核心（包括源代码）、XWindows、各种语言编译器、网络支持软件、各种服务器软件、Java 核心支持，以及中文平台、应用软件、游戏等。与直接从 Internet 上下载数百兆字节相比，发行版本其实更便宜，而且安装起来也更方便快捷。更为重要的是，商业软件公司能够为用户提供专业的技术支持，从而解决了用户的后顾之忧。

所有这些都为 Linux 操作系统的迅速发展奠定了坚实的基础。

1.1.3 Linux 在中国

最近几年来，我国的软件工作者也开始关注 Linux，并做了许多有益的工作。各种 Linux 中文化技术相继开发成功，Linux 的用户也在逐年增加。

由于系统软件由少数国外公司所垄断，因此我国的民族软件产业一直处于在“夹缝”中求生存的状况。以 Linux 为代表的、以开放源代码为特征的自由软件运动为我国民族软件产业的发展提供了一个很好的机会。

国家有关部门和许多致力于民族软件发展的公司看到了这个机会，并正在积极地利用这个机会，将我国软件产业的发展提高到一个新的阶段。正如前面所介绍的，许多国内软件公司正在积极地研究、探索 Linux 在我国的推广和普及。中国自由软件库（<http://freesoft.cei.gov.cn>）也已建成并投入使用。所有这些都为 Linux 在我国的推广和普及奠定了良好的基础。

自由软件毫无保留地提供源代码，使我们可以高起点起步，有利于我国软件行业在较短的时间内彻底改变面貌。

1.2 Linux 的技术特点

虽然是 UNIX 类的操作系统，但 Linux 并不是某种 UNIX 的翻版，而是一种完全从基础代码开始重写的操作系统，并集中了 BSD UNIX 和 System V 等一些著名 UNIX 版本的优点。BSD UNIX 是美国加州大学伯克利分校于 1977 年开始开发的，是一种著名的并富有特色的 UNIX 版本，Linux(包括其它许多现代 UNIX 商业版本)的许多先进特色和程序都出自这种 UNIX。遗憾的是 90 年代初期，BSD UNIX 的开发就已经基本停止，虽然如此，人们仍然能够从 Internet 上下载这种免费的 UNIX。对 Linux 有着重要影响的另一个 UNIX 版本是 System V。UNIX 的鼻祖 AT&T 开发出了多种 UNIX 版本，如 System III、System IV 和 System V。Linux 使用了 System V Release 4(SVR4)的许多技术，如信号灯进程间通信技术等。

1.2.1 Linux 的特点

Linux 具备了现代 UNIX 操作系统的主要特征，其主要的特点有：

1. 符合 POSIX 1003.1 标准

POSIX 1003.1 标准定义了一个最小的 UNIX 操作系统接口，任何操作系统只有符合这一标准，才有可能运行 UNIX 程序。考虑到 UNIX 具有丰富的应用程序，当今绝大多数操作系统都把符合 POSIX 1003.1 标准作为实现目标，Linux 也不例外，它完全支持 POSIX 1003.1 标准。另外，为了使 BSD 和 System V UNIX 上的程序能直接在 Linux 上运行(源代码级兼容，部分程序在二进制级兼容)，Linux 还增加了部分 System V 和 BSD 的系统接口，成为一个较完善的 UNIX 程序开发平台。

2. 支持多用户、多任务

Linux 是一个真正的多用户、多任务操作系统，它允许多个用户同时访问系统。另外，Linux 还支持真正的多任务，一个用户可以创建多个进程，并使各个进程协同工作来完成用户的需求。

3. 采用虚拟内存管理技术

Linux 支持请求页式虚拟内存管理技术，也意味着只有当前运行最需要的代码和数据才会装入到系统的物理内存。为了进一步优化内存的使用，Linux 还支持内存缓冲机制，空闲的内存可用于磁盘和设备缓存，从而加速了对代码和数据的访问，并能根据内存的使用情况自动对缓存的大小进行调整。

Linux 利用交换技术，为应用程序提供比实际物理内存大得多的虚拟内存。