

# Linux技术内幕



[Linux技术内幕 下载链接1](#)

著者:罗秋明

出版者:清华大学出版社

出版时间:2017-1-1

装帧:平装

isbn:9787302451006

本书内容分成两篇，第一篇是基本框架，第二篇是文件系统及相关内容。在篇的各章中：第1章先对Linux进行简要介绍并给出本书分析时所用的模型；第2章从C语言程序、可执行文件到进程的虚存空间影像的全过程作为起点，给读者建立起进程用户空间管理的概念；然后第3章讨论物理页帧如何支撑这些虚存空间，并且讨论了与物理空间一致的内核空间的管理；接着第4章就是进程的概念、进程的组织、进程切换和进程的创建撤销等活动；第5章专门讨论进程调度和负载均衡问题；后面4章继续讨论进程间通信、系统调用、内核的并发活动和同步问题。第二篇开始讨论盘根错节的文件系统：先在第10章分析文件系统和VFS的基本概念；然后在第11章讨论页高速缓存及块缓存；第12章分析了VFS的通用文件访问操作；第13章讨论ext2文件系统的具体格式和操作细节；接着第14

章讨论同步；第15章讨论内存回收问题；后第16章和第17章讨论设备管理和块设备问题。

本书以内存模型和时空模型为主要参考来分析各章的相关内容，给出了比较直观的图示，这不仅对初学者非常有用，对希望了解Linux内核的读者和相关开发人员也非常有参考价值。

作者介绍:

罗秋明，男，1974年生。深圳大学计算机与软件学院副教授、硕士生导师，国家高性能计算中心深圳分中心科研人员。

目录: 第1章 Linux内核概述 3

1.1 UNIX与Linux 3

1.1.1 UNIX 3

1.1.2 Linux 3

1.1.3 宏内核与微内核 6

1.1.4 Linux内核源码及版本 6

1.2 Linux内核模型 11

1.2.1 多视角下的内核 11

1.2.2 功能模型 14

1.2.3 内存模型 15

1.2.4 时空模型 16

1.2.5 特权模型 17

1.3 本书局限性 18

小结 18

第2章 进程影像 20

2.1 从源代码到进程 20

2.1.1 源代码、目标文件 20

2.1.2 可执行文件与进程影像 26

2.2 proc中的进程 32

2.2.1 进程内存空间 32

2.2.2 进程运行状态等信息 33

2.3 进程空间 35

2.3.1 进程空间描述符 36

2.3.2 虚存区域VMA 40

2.3.3 VMA属性 45

2.4 ELF可执行文件装入过程 48

2.4.1 ELF装入函数 49

2.4.2 ELF格式 49

2.5 进程空间的动态变化 54

2.5.1 VMA上的操作 54

2.5.2 文件映射 55

2.5.3 堆的调整 58

2.5.4 栈的变化 60

2.6 并发的进程空间 60

小结 61

第3章 虚拟空间的物理支撑 63

3.1 物理内存组织与管理 64

3.1.1 节点与内存域 64

3.1.2 物理页帧 79

3.1.3 buddy系统 84

3.1.4 页帧迁移 94

3.1.5 内存热插拔 98

3.2 地址映射与页表 99

3.2.1 分页机制与页表 99

3.2.2 缺页异常 103

3.3 内核空间 104

3.3.1 一致映射与高端内存 104

3.3.2 一致内存分配 108

3.3.3 非一致内存分配 109

3.3.4 slab分配器 115

小结 130

第4章 进程组织与基础行为 131

4.1 进程组织管理 131

4.1.1 PCB进程控制块 131

4.1.2 命名空间 135

4.1.3 进程标识 141

4.1.4 进程间关系 148

4.1.5 进程资源限制 151

4.2 进程创建与撤销 152

4.2.1 进程创建 152

4.2.2 execve系统调用 159

4.2.3 内核线程 160

4.2.4 Linux进程树 162

4.2.5 进程的撤销 167

4.3 进程切换 167

4.3.1 切换时机 168

4.3.2 切换过程 169

4.3.3 切换示例 178

小结 181

第5章 进程调度与负载均衡 193

5.1 调度与均衡基本框架 193

5.2 进程状态与转换 194

5.2.1 进程调度状态 194

5.2.2 进程状态变迁 196

## 5.3 进程调度 198

### 5.3.1 调度框架 198

### 5.3.2 完全公平调度 215

### 5.3.3 实时调度 230

### 5.3.4 STOP和IDLE调度类 234

### 5.3.5 调度控制与proc接口 236

## 5.4 负载均衡 241

### 5.4.1 处理器层次结构 242

### 5.4.2 调度的层次管理 243

### 5.4.3 CFS任务的负载均衡 250

### 5.4.4 实时负载均衡 254

## 小结 258

## 第6章 进程间通信与同步 259

### 6.1 管道通信 260

#### 6.1.1 无名管道 260

#### 6.1.2 命名管道 261

#### 6.1.3 管道数据结构 261

#### 6.1.4 管道操作 266

### 6.2 System V IPC 269

#### 6.2.1 IPC标识与命名空间 269

#### 6.2.2 IPC公共框架 275

#### 6.2.3 IPC信号量 279

#### 6.2.4 IPC消息队列 283

#### 6.2.5 IPC共享内存 287

### 6.3 信号 290

#### 6.3.1 信号分类 290

#### 6.3.2 数据结构 294

6.3.3 信号产生与发送 302

6.3.4 信号的递交和处理 305

小结 312

第7章 内核活动 313

7.1 中断分类 313

7.1.1 x86的中断和异常 314

7.1.2 后半部机制与软中断 315

7.1.3 中断相关概念的关系 315

7.2 中断处理 316

7.2.1 中断号 317

7.2.2 中断描述符表 320

7.2.3 公共入口 325

7.2.4 异常处理 329

7.3 高层中断处理 332

7.3.1 转向高层处理 333

7.3.2 中断的高层数据结构 336

7.3.3 中断返回处理 342

7.3.4 中断的线程化 349

7.4 中断嵌套与中断管理 350

7.4.1 中断嵌套与中断上下文 350

7.4.2 中断管理 353

7.5 软中断和tasklet 355

7.5.1 中断的下半部 355

7.5.2 软中断执行时机 356

7.5.3 相关数据结构 360

7.5.4 软中断的执行 361

7.5.5 软中断的相关操作 363

7.5.6 tasklet 365

7.6 工作队列 369

7.6.1 工作队列机制 369

7.6.2 cmwq数据结构 371

7.6.3 工作项 371

7.6.4 cmwq工作队列 375

7.6.5 工作者池worker\_pool 383

7.6.6 并发度、应急处理等 386

7.7 系统调用 388

7.7.1 POSIX API、C库和系统调用 388

7.7.2 系统调用的实现 389

小结 399

第8章 时间管理 400

8.1 时间管理框架 400

8.1.1 基本概念 400

8.1.2 时间中断和事件 406

8.1.3 clock\_event\_device与tick\_device 407

8.1.4 TIMER\_SOFTIRQ软中断 413

8.1.5 timekeeper 414

8.2 定时器 416

8.2.1 低分辨率定时器 416

8.2.2 高精度定时器 419

8.2.3 模拟tick事件 422

8.2.4 通知链技术 423

小结 424

第9章 内核并发与同步 426

9.1 同步的需求 426

9.1.1 内核并发情形	426
9.1.2 内核抢占	429
9.2 内核共享变量的保护	432
9.2.1 被保护对象	432
9.2.2 保护原则	433
9.2.3 禁止内核并发	435
9.3 内核同步手段	437
9.3.1 原子操作	437
9.3.2 自旋锁、读写锁和顺序锁	439
9.3.3 RCU机制	444
9.3.4 顺序和屏障	447
9.3.5 信号量与互斥量	448
9.3.6 等待队列与完成变量	452
9.3.7 每CPU变量	455
小结	458

## 第二篇 盘根错节的文件系统

第10章 文件系统	461
10.1 文件系统的抽象层次	461
10.1.1 进程视角下的文件	462
10.1.2 VFS虚拟文件系统	468
10.1.3 多角度分层模型	472
10.2 VFS核心对象	475
10.2.1 文件对象	475
10.2.2 目录项对象	479
10.2.3 索引节点对象	484
10.2.4 超级块对象	490



10.3 文件系统类型与挂载 495

10.3.1 文件系统类型与注册 495

10.3.2 挂载操作 503

10.3.3 挂载模式 513

10.3.4 特殊文件系统 519

小结 525

第11章 页缓存和块缓存 526

11.1 页高速缓存 527

11.1.1 address\_space 528

11.1.2 页高速缓存的组织 530

11.1.3 反向映射 534

11.2 块高速缓存 540

11.2.1 块缓存 540

11.2.2 LRU块缓存 542

11.2.3 块缓存操作 543

小结 547

第12章 VFS的文件操作 548

12.1 VFS系统调用 548

12.2 open()与close()系统调用 549

12.2.1 open的框架 549

12.2.2 文件定位过程 552

12.2.3 close()系统调用 558

12.3 读/写系统调用 558

12.3.1 入口代码 560

12.3.2 通用write写例程 568

12.3.3 通用read读例程 570

12.3.4 其他读写细节 583

12.3.5 向BIO层提交请求 587

小结 588

第13章 ext2文件系统 590

13.1 ext2磁盘数据结构 590

13.1.1 磁盘分区的组织 590

13.1.2 块组描述符和位图 591

13.1.3 盘上和内存数据结构 592

13.2 ext2超级块 593

13.2.1 ext2超级块数据结构 593

13.2.2 挂载与访问 597

13.3 ext2索引节点 598

13.3.1 盘上ext2索引节点 599

13.3.2 内存ext2索引节点 601

13.3.3 inode\_operations 603

13.3.4 ext2地址空间与文件操作 604

13.4 目录及目录项 607

13.4.1 ext2\_dir\_entry 607

13.4.2 ext2\_lookup() 609

小结 610

第14章 页缓存同步（回写） 611

14.1 同步/回写、交换与回收 611

14.2 脏页同步（回写） 613

14.2.1 回写机制演变 613

14.2.2 同步时机与框架 615

14.2.3 基本数据结构 617

14.3 回写接口 627

14.3.1 sync系列系统调用 627

- 14.3.2 sys\_sync() 628
- 14.3.3 sys\_syncfs 633
- 14.3.4 单个文件的同步 635
- 14.3.5 被动回写 637
- 14.4 回写工作队列 638
  - 14.4.1 初始化 638
  - 14.4.2 工作队列处理函数 640
- 14.5 回写操作 643
  - 14.5.1 do\_writepages() 644
  - 14.5.2 ext2\_writepages() 645
  - 14.5.3 回写等待 647
- 小结 649
- 第15章 内存回收与交换 650
  - 15.1 页帧回收 650
    - 15.1.1 直接释放 650
    - 15.1.2 LRU页帧组织 651
    - 15.1.3 PFRA回收算法 655
  - 15.2 核心回收操作 659
    - 15.2.1 shrink\_zone() 659
    - 15.2.2 shrink\_slab() 666
    - 15.2.3 解除页表映射 667
  - 15.3 交换 667
    - 15.3.1 交换功能 668
    - 15.3.2 交换分区 668
    - 15.3.3 交换缓存 673
  - 小结 675
- 第16章 设备管理 676

16.1 设备管理组织	676
16.1.1 设备驱动模型	677
16.1.2 sysfs	678
16.1.3 基础组件	679
16.1.4 容器	682
16.2 设备的VFS接口	688
16.2.1 设备文件	689
16.2.2 从VFS中访问设备	691
16.3 字符设备	693
16.3.1 设备的散列组织	694
16.3.2 初始化与注册	695
16.3.3 打开字符设备	696
16.4 PCI设备	698
16.4.1 pci_bus_type和pci_bus	699
16.4.2 pci_driver	700
16.4.3 pci_dev	701
16.4.4 uevent	704
小结	705
第17章 块设备	706
17.1 基本概念	706
17.1.1 块设备层	706
17.1.2 传送单位	708
17.2 块设备层组件	709
17.2.1 磁盘与磁盘分区	709
17.2.2 块设备	712
17.2.3 请求队列	715
17.3 提交请求及处理	728

17.3.1 plug/unplug机制	728
17.3.2 提交请求	733
17.3.3 提交到驱动程序	742
17.3.4 硬盘的request_fn	744
17.3.5 中断处理	746
17.4 IO调度	749
17.4.1 IO调度器	749
17.4.2 调度器数据结构	751
17.5 初始化及注册	754
17.5.1 块设备初始化	754
17.5.2 硬盘初始化	759
小结	764
附录	765
后记	774
• • • • •	( <a href="#">收起</a> )

[Linux技术内幕\\_下载链接1](#)

## 标签

Linux

操作系统

linux内核导读

计算机

kernel

技术

2017

计算机科学

## 评论

本书是基于3.13版本内核来写的，但是这个版本的address\_space中的i\_mmap成员，还是使用的红黑树类型，但是在11.1.3这一节全篇使用优先树来描述该结构，我不知道作者是不是没有认真看这个版本的代码？而在本章的最开始，又是以红黑树来描述的。本来对本书的评价和观感还不错，看到这段重大BUG开始有点寒心，这不是细心与否能解释的过去的了。

-----  
有不足,但有可取之处,内容还算比较新;颇有收获.  
部分图例画的很好,便于理解;核心的子系统都有讲。有明显的参考其他内核书籍的痕迹。缺点思想深度不够,以及自己发明的突兀术语。

-----  
粗读完一遍,基本理解了核心骨架,具体细节性,策略性的内容需要更深入的围绕着源码分析来理解。书值得推荐的,读的过程中能感受到作者的用心。虽然有少许错误,但鉴于Linux本身的复杂,不要期待从一本书或简单的几本书中就能彻底理解清楚,本身这是迭代的过程。

-----  
给4星吧,比《linux内核源代码情景分析》还是差一些,但是确实还是有一些可取的地方,比如对可执行文件的section和fragment的区别,之前看《程序员的自我修养》没懂,看这本书明白了,还有就是物理内存管理,之前对NUMA的物理内存结构不理解,也是这本书看懂了的,还是有一些收获的。

-----  
[Linux技术内幕 下载链接1](#)

## 书评

-----  
[Linux技术内幕\\_下载链接1](#)