

# THE **LINUX** PROGRAMMING INTERFACE

A Linux and UNIX® System Programming Handbook

MICHAEL KERRISK



# THE LINUX PROGRAMMING INTERFACE

A Linux and UNIX System Programming Handbook

MICHAEL KERRISK

no starch press

San Francisco

*Translated by: Kevin Zhang*

2011-01-26 – 2011-05-15(TODO)

**本资料仅供学习所用，请于下载后 24 小时内删除，否则引起的任何后果均由您自己承担。本书版权归原作者所有，如果您喜欢本书，请购买正版支持作者。**

## 目录

前言.....	7
主题 .....	7
目标读者.....	7
Linux 和 UNIX .....	8
使用和组织.....	8
例子程序.....	9
练习 .....	10
标准和可移植性 .....	10
Linux 内核和 C 库版本 .....	11
其它语言使用编程接口 .....	11
关于作者.....	11
致谢 .....	11
许可 .....	12
网站和例子程序源代码.....	12
反馈 .....	12
第 1 章 历史和标准 .....	13
第 2 章 基础概念 .....	14
第 3 章 系统编程概念 .....	15
第 4 章 文件 I/O: 统一的 I/O 模型 .....	16
第 5 章 文件 I/O: 更多细节.....	17
第 6 章 进程 .....	18
第 7 章 内存分配.....	19
第 8 章 用户和组.....	20
第 9 章 进程凭证.....	21
第 10 章 时间 .....	22
第 11 章 系统限制和选项.....	23
第 12 章 系统和进程信息.....	24

第 13 章 文件 I/O 缓冲 .....	25
第 14 章 文件系统 .....	26
第 15 章 文件属性 .....	27
第 16 章 扩展属性 .....	28
第 17 章 访问控制列表 .....	29
第 18 章 目录和链接 .....	30
第 19 章 监控文件事件 .....	31
第 20 章 信号：基础概念 .....	32
第 21 章 信号：信号处理器 .....	33
第 22 章 信号：高级特性 .....	34
第 23 章 定时器和睡眠 .....	35
第 24 章 进程创建 .....	36
第 25 章 进程结束 .....	37
第 26 章 监控子进程 .....	38
第 27 章 程序执行 .....	39
第 28 章 进程创建和程序执行的更多细节 .....	40
第 29 章 线程：介绍 .....	41
第 30 章 线程：同步 .....	42
第 31 章 线程：线程安全和线程存储 .....	43
第 32 章 线程：线程取消 .....	44
第 33 章 线程：更多细节 .....	45
第 34 章 进程组、会话和任务控制 .....	46
第 35 章 进程优先级和调度 .....	47
第 36 章 进程资源 .....	48
第 37 章 Daemon .....	49
第 38 章 编写安全的特权程序 .....	50
第 39 章 能力 .....	51
第 40 章 登录会计 .....	52

第 41 章 共享库基础 .....	53
第 42 章 共享库高级特性 .....	54
第 43 章 进程间通信简介 .....	55
第 44 章 管道和 FIFO .....	56
第 45 章 System V IPC 介绍 .....	57
第 46 章 System V 消息队列 .....	58
第 47 章 System V 信号量 .....	59
第 48 章 System V 共享内存 .....	60
第 49 章 内存映射 .....	61
第 50 章 虚拟内存操作 .....	62
第 51 章 POSIX IPC 介绍 .....	63
第 52 章 POSIX 消息队列 .....	64
第 53 章 POSIX 信号量 .....	65
第 54 章 POSIX 共享内存 .....	66
第 55 章 文件锁 .....	67
第 56 章 Sockets: 介绍 .....	68
第 57 章 Sockets: UNIX Domain .....	69
第 58 章 Sockets: TCP/IP 网络基础 .....	70
第 59 章 Sockets: Internet Domain .....	71
第 60 章 Sockets: 服务器设计 .....	72
第 61 章 Sockets: 高级主题 .....	73
第 62 章 终端 .....	74
第 63 章 可选 I/O 模型 .....	75
第 64 章 伪终端 .....	76
附录 A: 跟踪系统调用 .....	77
附录 B: 解析命令行参数 .....	78
附录 C: 转换 NULL 指针 .....	79
附录 D: 内核配置 .....	80

附录 E：更多信息来源 .....	81
附录 F：部分习题解答.....	82
参考书目 .....	83
索引.....	84

# 前言

## 主题

本书描述 Linux 编程接口——Linux（UNIX 操作系统的一种免费实现）提供的系统调用、库函数、和其它底层接口。这些接口被直接或间接地使用在 Linux 上运行的每个程序中。它们允许应用程序完成各种任务：如文件 I/O、创建删除文件和目录、创建新进程、执行程序、设置定时器、本机进程和线程间通信、通过网络连接的不同机器进程间通信等等。这些底层接口有时候也叫做系统编程接口。

尽管本书关注于 Linux，但我也非常注意标准和可移植性问题，清晰地区分了 Linux 特有的接口、多数 UNIX 实现共有的特性、以及 POSIX 和 Single UNIX Specification 标准定义的特性。因此本书也提供了 UNIX/POSIX 编程接口的详尽描述，能够适用于编写 UNIX 系统应用或跨平台应用的程序员。

## 目标读者

本书主要面向以下读者：

- 为 Linux、UNIX、或者其它遵循 POSIX 的系统开发应用的程序员和软件设计师；
- 在 Linux、UNIX、或其它操作系统之间移植应用的程序员；
- Linux 或 UNIX 系统编程课程的教师和高年级学生；
- 希望深入理解 Linux/UNIX 编程接口，以及系统软件是如何实现的系统管理员和“高级用户”。

我假设你拥有一定的编程经验，但不要求系统编程经验。我还假设你了解 C 编程语言，并且知道如何使用 shell 和常用的 Linux 或 UNIX 命令。如果你是 Linux/UNIX 的新手，你会发现第 2 章非常有用，我们以程序员的视角来讲述 Linux 和 UNIX 的基础概念。

## Linux 和 UNIX

本书原本可以纯粹地讲解标准 UNIX（也就是 POSIX）系统编程，因为 UNIX 和 Linux 的大多数特性都是相同的。不过虽然编写可移植程序是很好的目标，理解 Linux 对标准 UNIX 编程接口的扩展也是非常重要的。理由之一是 Linux 非常流行；其二是有时候为了性能、或使用标准 UNIX 没有的功能，我们不得不使用非标准的扩展（所有 UNIX 实现都提供类似的非标准扩展）。

因此本书在适用于标准 UNIX 的程序员时，还提供了 Linux 特定编程特性的详细描述。这些特性包括：

- `epoll`，获得文件 I/O 事件通知的机制；
- `inotify`，监控文件和目录改变的机制；
- 能力，授予进程一组超级用户能力的机制；
- 扩展属性；
- `i-node` 标志；
- `clone()` 系统调用；
- `/proc` 文件系统
- Linux 对文件 I/O、信号、定时器、线程、共享库、进程间通信、和 `socket` 的特殊实现细节。

## 使用和组织

你至少可以按两种方式使用本书：

- 作为 Linux/UNIX 编程接口的介绍手册。你可以从头到尾阅读本书。后续章节建立在之前章节的基础之上，我尽量避免依赖后续章节的情况。
- 作为 Linux/UNIX 编程接口的索引参考手册。详细的索引和频繁的交叉引用，允许你随机地阅读任何主题。

我把本书分为以下几部分：

1. 背景和概念：UNIX、C 和 Linux 的历史；UNIX 标准简介（第 1 章）；以程



序员的视角介绍 Linux 和 UNIX 的基本概念（第 2 章）；Linux 和 UNIX 系统编程的基本概念（第 3 章）。

2. 系统编程接口的基础特性：文件 I/O（第 4 章和第 5 章）；进程（第 6 章）；内存分配（第 7 章）；用户和组（第 8 章）；进程凭证（第 9 章）；定时器（第 10 章）；系统限制和选项（第 11 章）；获取系统和进程信息（第 12 章）。
3. 系统编程接口的高级特性：文件 I/O 缓冲（第 13 章）；文件系统（第 14 章）；文件属性（第 15 章）；扩展属性（第 16 章）；访问控制列表（第 17 章）；目录和链接（第 18 章）；监控文件事件（第 19 章）；信号（第 20 章到第 22 章）；定时器（第 23 章）。
4. 进程、程序、和线程：进程创建、进程结束、监控子进程、执行程序（第 24 章到第 28 章）；POSIX 线程（第 29 章到第 33 章）。
5. 进程和程序的高级主题：进程组、会话、任务控制（第 34 章）；进程优先级和调度（第 35 章）；进程资源（第 36 章）；daemon（第 37 章）；编写安全的特权程序（第 38 章）；能力（第 39 章）；登录会计（第 40 章）；共享库（第 41 章到第 42 章）。
6. 进程间通信（IPC）：IPC 简介（第 43 章）；管道和 FIFO（第 44 章）；System V IPC——消息队列、信号量、共享内存（第 45 章到第 48 章）；内存映射（第 49 章）；虚拟内存操作（第 50 章）；POSIX IPC——消息队列、信号量、共享内存（第 51 章到第 54 章）；文件锁（第 55 章）。
7. Socket 和网络编程：IPC 和 socket 网络编程（第 56 章到第 61 章）。
8. 高级 I/O 主题：终端（第 62 章）；可选 I/O 模型（第 63 章）；伪终端（第 64 章）。

## 例子程序

我用短小但完整的例子程序来阐述多数接口的使用方法，这些例子都被设计为很容易就能从命令行体验，来查看不同的系统调用和库函数如何工作。所以本书包含大量的示例代码——大概 15000 行 C 代码和 shell 会话日志。

尽管阅读和试验例子程序是不错的起点，掌握本书讨论的概念最有效的方法是编写代码，按你的想法修改例子程序，或者编写新程序都可以。

本书的所有源代码都可以在网站上下载。源代码包含许多书中没有的程序。这些程序的目的和细节在注释中都有相关描述。我提供了 **Makefile** 编译这些程序，以及一个 **README** 文件，给出了例子程序更多的细节信息。

源代码采用 **GNU Affero** 通用公共授权版本 3，可以自由分发和修改。源代码中也包含一份该协议的拷贝。

## 练习

多数章节都以一组练习结束，其中一些是要你按不同方式来试验例子程序，另外一些是该章讨论过的概念相关的问题，还有就是要求你来编写代码以巩固你对本书的理解。你可以在附录 F 找到部分练习的解答。

## 标准和可移植性

贯穿整本书，我都对可移植性问题特别地关注。你会发现很多相关标准的引用，特别是 **POSIX.1-2001** 和 **Single UNIX 规范版本 3 (SUSv3)** 标准。同时你还将看到这些标准最新修订的细节改变，也就是 **POSIX.1-2008** 和 **SUSv4** 标准。（由于 **SUSv3** 是更大的修订版本，也是本书编写时最广泛有效的 **UNIX** 标准，本书讨论的标准大多是 **SUSv3**，并标注出 **SUSv4** 不同的地方。除非我明确地提到，你可以假设我们对 **SUSv3** 规范的描述也适用于 **SUSv4**）。

对于那些不是标准的特性，我会指出在不同 **UNIX** 实现间的差别。我还会突出那些 **Linux** 特定的特性，以及 **Linux** 与其它 **UNIX** 对系统调用和库函数实现上的细小差别。当某个特性我没有明确指出是 **Linux** 专有时，你也通常可以假设它在多数或所有 **UNIX** 上都有实现。

本书大多数例子程序我都在 **Solaris**、**FreeBSD**、**Mac OS X**、**Tru64 UNIX**、和 **HP-UX** 上测试通过（除了那些 **Linux** 独有的特性）。为了提高代码在这些系统上的可移植性，本书网站上提供的某些例子程序有一些额外的代码。

## Linux 内核和 C 库版本

本书主要关注 Linux 2.6.x 系列,这是本书写作时最广泛使用的内核版本。Linux 2.4 的某些细节也会提到,我也会指出 Linux 2.4 和 2.6 的区别。当 Linux 2.6.x 系列出现了新特性时(例如 2.6.34),我也会特别指出相应的内核版本号。

至于 C 库,本书则主要关注于 GNU C 库(glibc)版本 2。当然,glibc 2.x 系列版本存在差异时,我也会特别指出。

在本书即将印刷时,Linux 内核刚刚发布了 2.6.35 版本,glibc 则已经发布 2.12 版本。本书完全适用于这两个软件版本。Linux 内核和 glibc 将来接口的变化,会在本书的网站上列出。

## 其它语言使用编程接口

尽管例子程序用 C 语言编写,你也可以在其它编程语言中使用本书讨论的接口——例如编译型语言 C++、Pascal、Modula、Ada、FORTRAN、D; 解释型语言 Perl、Python、Ruby 等。(Java 则需要采用一种不同的方式 JNI)。不同的语言要获取必要的常量定义和函数声明,需要使用不同的技术(C++除外),另外传递函数参数时可能也需要一点额外的工作。此外就没有太大的区别了,核心概念其实都是一样的。因此即使你使用其它的编程语言,你也会发现本书提供的信息是适用的。

## 关于作者

(略)

## 致谢

(略)

## 许可

电子工程学会和开放组织非常友好地许可我引用 IEEE Std 1003.1, 2004 版本，以及信息技术标准——可移植操作系统接口(POSIX)，开放组织基本规范 Issue6。完整的标准可以在 <http://www.unix.org/version3/online.html> 上在线查阅。

## 网站和例子程序源代码

你可以在 <http://www.man7.org/tlpi> 上找到关于本书更多的信息，包括勘误表和例子程序的源代码。

## 反馈

我非常欢迎代码 bug 报告、代码改进建议、以及代码可移植性的提高。同样我也欢迎本书的 bug 报告和改进建议。由于 Linux 编程接口总是在变化，我也非常高兴能获得关于本书将来版本的改进意见，包括新特性和变化特性。

Michael Timothy Kerrisk

Munich, Germany and Christchurch, New Zealand

August 2010

[mtk@man7.org](mailto:mtk@man7.org)

# 第 1 章 历史和标准

Linux 是 UNIX 操作系统家族的成员之一。在计算机的术语里，UNIX 已经拥有很悠久的历史。第 1 章的前半部分简述 UNIX 的历史。我们首先描述 UNIX 系统和 C 编程语言的起源，然后讲述导致 Linux 发展成为今天这个样子的两个关键因素：GNU 项目和 Linux 内核的开发。

UNIX 系统最显著的特点之一是它的开发不是被一个厂商或组织控制。相反许多商业和非商业组织都为 UNIX 的发展做出了贡献。UNIX 也因此增加了许多革新的特性，但同时也导致 UNIX 各个实现之间的分歧越来越大，编写一个能运行于所有 UNIX 实现的应用也变得非常困难。于是产生了 UNIX 的标准化运动，我们将在本章后半部分进行讨论。

## 1.1 UNIX 和 C 简史

## 第 2 章 基础概念

## 第 3 章 系统编程概念

## 第 4 章 文件 I/O: 统一的 I/O 模型



## 第 5 章 文件 I/O: 更多细节

## 第 6 章 进程

## 第 7 章 内存分配

## 第 8 章 用户和组

## 第 9 章 进程凭证

## 第 10 章 时间

## 第 11 章 系统限制和选项

## 第 12 章 系统和进程信息



## 第 13 章 文件 I/O 缓冲

## 第 14 章 文件系统

## 第 15 章 文件属性

## 第 16 章 扩展属性

## 第 17 章 访问控制列表

## 第 18 章 目录和链接

## 第 19 章 监控文件事件

## 第 20 章 信号：基础概念



## 第 21 章 信号：信号处理器

## 第 22 章 信号：高级特性

## 第 23 章 定时器和睡眠

## 第 24 章 进程创建

## 第 25 章 进程结束

## 第 26 章 监控子进程

## 第 27 章 程序执行

## 第 28 章 进程创建和程序执行的更多细节



## 第 29 章 线程：介绍

## 第 30 章 线程：同步

## 第 31 章 线程：线程安全和线程存储

## 第 32 章 线程：线程取消

## 第 33 章 线程：更多细节

## 第 34 章 进程组、会话和任务控制

## 第 35 章 进程优先级和调度

## 第 36 章 进程资源



## 第 37 章 Daemon

## 第 38 章 编写安全的特权程序

## 第 39 章 能力

## 第 40 章 登录会计

## 第 41 章 共享库基础

## 第 42 章 共享库高级特性

## 第 43 章 进程间通信简介

## 第 44 章 管道和 FIFO



## 第 45 章 System V IPC 介绍

## 第 46 章 System V 消息队列

## 第 47 章 System V 信号量

## 第 48 章 System V 共享内存

## 第 49 章 内存映射

## 第 50 章 虚拟内存操作

## 第 51 章 POSIX IPC 介绍

## 第 52 章 POSIX 消息队列



## 第 53 章 POSIX 信号量

## 第 54 章 POSIX 共享内存

## 第 55 章 文件锁

## 第 56 章 Sockets: 介绍

## 第 57 章 Sockets: UNIX Domain

## 第 58 章 Sockets: TCP/IP 网络基础

## 第 59 章 Sockets: Internet Domain

## 第 60 章 Sockets: 服务器设计



## 第 61 章 Sockets: 高级主题

## 第 62 章 终端

## 第 63 章 可选 I/O 模型

## 第 64 章 伪终端

## 附录 A: 跟踪系统调用

## 附录 B：解析命令行参数

## 附录 C: 转换 NULL 指针

## 附录 D：内核配置



## 附录 E：更多信息来源

## 附录 F： 部分习题解答

## 参考书目

# 索引