EMBEDDED SYSTEM

嵌入式技术与应用丛书

Linux内核分析及高级编程

吴国伟　李 张　任广臣　编著

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

TP316.81/150
2008

嵌入式技术与应用丛书

Linux 内核分析及高级编程

吴国伟 李 张 任广臣 编著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

内 容 简 介

本书以 Linux 2.6 内核分析和编程应用来讲解高级操作系统设计和应用方面的问题。本书是编著者在多年的教学和科研的基础上编写的，在结构上每章先进行内核设计的分析，然后是实践案例的设计，最后是情景分析应用。全书共有 13 章，主要包括 Linux 内核中进程管理、进程间同步通信、内存管理、文件系统、I/O 设备管理、模块机制、定时器机制，以及利用内核机制进行高级编程等。

本书可作为计算机科学与技术、软件工程或相关专业的高级操作系统教材，也可供从事嵌入式软件和系统软件设计等的科技人员自学或参考。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目（CIP）数据

Linux 内核分析及高级编程/吴国伟，李张，任广臣编著.—北京：电子工业出版社，2008.1
（嵌入式技术与应用丛书）
ISBN 978-7-121-05244-6

Ⅰ. L… Ⅱ. ①吴…②李…③任… Ⅲ. Linux 操作系统—程序设计 Ⅳ. TP316.89

中国版本图书馆 CIP 数据核字（2007）第 164455 号

责任编辑：高买花　　特约编辑：陈宁辉
印　　刷：北京京科印刷有限公司
装　　订：
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×1 092　1/16　印张：17.75　字数：454 千字
印　　次：2008 年 1 月第 1 次印刷
定　　价：29.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

服务热线：（010）88258888。

出 版 说 明

嵌入式技术是21世纪最具生命力的新技术之一，经过近几年的快速发展，已经成为电子信息产业中最具增长力的一个分支。随着手机、掌上电脑、GPS、机顶盒等新兴产品的大量应用，嵌入式系统的设计正成为软硬件工程师越来越关心的话题。面对不断涌现的技术需求和发展机遇，各大嵌入式系统开发商、各科研院所的研发人员都急需一套全方位、针对性强，且具有实际指导意义的嵌入式技术类书籍；各高等院校相关专业的本科生、研究生也迫切希望了解、掌握嵌入式系统的开发技巧，以推动嵌入式技术在各领域的广泛应用和快速发展。

《嵌入式技术与应用丛书》正是针对当前技术与市场需求，由国内站在IT业前沿并有实践开发经验的嵌入式系统专家，以实用技术为主线，理论联系实际，将他们在理论研究与实践工作中积累的大量经验和体会有机地融于一体，以丛书的形式奉献给广大读者！

本丛书由基础理论类、硬件设计类、软件开发类、综合应用类书籍组成，立足当前嵌入式技术的发展趋势、核心技术及其主要应用领域，将技术热点与实践应用紧密结合，以实际应用为主线，融合关键性嵌入式设计技术，围绕嵌入式设计理论、开发流程、嵌入式软件验证及测试、代码可重构以及代码优化等方面进行深入浅出的讲解和论述。

读者群定位于高等院校相关领域的高年级学生，科研、开发人员，嵌入式相关领域设计人员等，本丛书可作为嵌入式领域学习、开发人员的参考资料，也可作为高等院校相关专业师生的教学参考书。

本丛书的出版得到了业界许多专家、学者的鼎力相助，对此表示衷心的感谢！同时，热切欢迎广大读者提出宝贵意见，或者推荐更多优秀选题（gmholife@hotmail.com），共同为嵌入式技术的发展添砖加瓦！

电子工业出版社

2007年6月

前　言

Linux 操作系统开发源代码，目前已被广泛应用于通用 PC、服务器、嵌入式设备上。在讲授研究生的高级操作系统课程时，一直在寻找适合研究生教育，能提高研究生的分析能力和创新能力的高级操作系统的教材，但很少见这样的教材。经过六年的教学和科研积累，我们编写了这本《Linux 内核分析及高级编程》，以 Linux 内核最新版本 2.6.20 的分析和高级应用为主要内容，目的在于通过本书的训练，能让学生提高系统软件的分析和设计能力。本书的编写思想是首先带领读者分析内核关键部分的代码，提炼出关键的数据结构和算法，给出处理流程，然后给出应用范例，让读者知道如何去应用，最后给出应用场景，让读者思考如何用学到的知识去解决这个场景。每一章还给出了本章的知识点，并为读者留出了思考题。本书意在通过“庖丁解牛”，让读者熟悉 Linux 内核，继而达到灵活运用的目的。

本书的章节安排：

第 1 章：Linux 内核简介，让读者对 Linux 内核有一个基本的认识和了解。

第 2 章：Linux 内核探索工具，让读者掌握对内核进行探索所必需的工具。

第 3 章：Linux 内核引导分析，首先从引导部分分析引导的过程和机制。

第 4 章：进程管理分析，分析 Linux 对进程的管理机制，包括关键的数据结构和处理流程、进程的创建方法等。

第 5 章：调度分析，分析 Linux 进程调度的算法。

第 6 章：系统调用，分析 Linux 系统调用的实现机制，并学会如何设计自己的系统调用。

第 7 章：中断机制，分析 Linux 下中断的处理机制，学会设计自己的中断。

第 8 章：内核同步机制，分析 Linux 内核的同步机制，包括锁等，学会使用同步机制完成场景的设计。

第 9 章：定时器和时间管理机制。

第 10 章：模块机制，分析 Linux 内核模块的实现方法，学会设计和添加新的模块。

第 11 章：内存管理机制，分析 Linux 下对内存的管理方法。

第 12 章：虚拟文件系统。

第 13 章：I/O 设备管理，分析 Linux 下对设备的管理方法，学会典型设备的驱动设计方法，包括字符设备、块设备等。

在本书编写过程中，研究生张杰、赵修伟、范庆娜、徐子川做了很多编辑和源程序开发的工作，在此表示感谢。

本书可作为计算机科学与技术、软件工程或相关专业的高级操作系统教材，也可供从事嵌入式软件和系统软件设计等的科技人员自学或参考。

编著者

2007 年 9 月

此为试读，需要完整PDF请访问：www.ertongbook.com

目　录

此为试读,需要完整PDF请访问: www.ertongbook.com

第 1 章　Linux 内核简介

本章从整体上对 Linux 内核进行说明。从顺序上说，本章首先介绍 Linux 的起源及一些发行版本的特性，然后介绍内核版本信息和体系结构，以及 Linux 源代码的组成，最后介绍如何编译内核。

1.1　Linux 简介

1.1.1　Linux 发展历史

Linux 操作系统是于 1991 年，由芬兰赫尔辛基大学计算机科学系的一名学生 Linus Benedict Torvalds 首先独立创建的，后来经过全球的电脑爱好者、黑客、软件开发人员等的参与，使 Linux 逐渐成为一个完善的操作系统。1991 年 10 月 5 日，Linus 在 comp.os.minix 新闻组上发布消息，正式向外宣布 Linux 内核系统的诞生（Free minix-like kernel sources for 386-AT）。

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，ARM，DEC Alpha，SUN Sparc，M68000，MIPS 和 PowerPC 等多种计算机平台上，已经成为应用广泛、可靠性高、功能强大的计算机操作系统。

Linux 操作系统是 UNIX 操作系统的一个克隆版本。UNIX 操作系统是美国贝尔实验室的 Ken.Thompson 和 Dennis Ritchie 于 1969 年夏在 DEC PDP-7 小型计算机上开发的一个分时操作系统。当时使用的是 BCPL 语言（基本组合编程语言），后经 Dennis Ritchie 于 1972 年用移植性很强的 C 语言进行了改写，使得 UNIX 系统在大专院校得到了推广。但后来由于 UNIX 的商品化，其源代码受到了版权保护，这在一定程度上影响了 UNIX 的普及和推广。

再后来出于教学的需要，荷兰教授 Andrew S. Tanenbaum（AST）编写了一个小型的类 UNIX 操作系统 MINIX。AST 在荷兰 Amsterdam 的 Vrije 大学数学与计算机科学系统工作，是 ACM 和 IEEE 的资深会员（全世界也只有很少人是两会的资深会员），共发表了 100 多篇文章，5 本计算机书籍。MINIX 是他于 1987 年编制的，主要用于学生学习操作系统原理。到 1991 年时版本是 1.5。目前主要有两个版本在使用：1.5 版和 2.0 版，当时该操作系统在大学使用是免费的，但其他用途不是，当然目前都已经是免费的，可以从许多 FTP 站上下载。

作为一个操作系统，MINIX 并不是优秀者，它虽然是一个不错的教学工具，却缺乏实用价值，正是看到了 MINIX 的这个缺点，Linus 决定自己编写一个操作系统——Linux。他想以 MINIX 为起点，基本上按照 UNIX 的设计，并取各种版本的优点，在 PC 上实现并开发出一个真正实用的 UNIX 内核。

1.1.2　常见的 Linux 介绍

目前读者所能接触到的 Linux 主要有 Red Hat、Slackware、Debian、SuSE、OpenLinux、TurboLinux、Red Flag、Mandarke、BluePoint 等。

（1）Red Hat（http://www.redhat.com）

Red Hat 以容易安装著称，初学者安装这个版本，遇到挫折的机会几乎是零，如果您对安装 Windows 9x 已驾轻就熟的话，Red Hat Linux 的安装一定难不倒您。

Red Hat 另一个优点是它的 RPM（Red Hat Package Manager）。以往在安装软件时，最让使用者伤脑筋的是：软件在解开压缩前先要新建一个目录，然后将软件搬进去解压，解压后，有些部分可能需要搬到另一个目录中去，当要搬移的项目多时，做这些工作就是件苦差事了。而 RPM（包裹管理器）就针对这一点，能将所有要安装的路径全部安排好，当使用者解开有.rpm 扩展名的文件时，会将当初打包该文件时设定好的路径档案先检查一次，然后依照档案里的设定将各个文件解开，送它们到应该去的地方；不只如此，它还会制作安装记录，当使用者要移除其中任一个 RPM 文件时，系统会根据安装记录将该文件反安装，这种做法绝对准确，不会像 Windows 那样会移除不该拿掉的东西。

Red Hat Linux 可以说是相当成功的一个产品，Red Hat 公司有官方版本（official）供使用者购买，也提供了自由的 FTP 站供大众直接下载。官方版本与自由下载版本的差异在于，官方版本多提供了一些商用软件和印刷精美的说明书。

（2）Slackware（http://www.cdrom.com）

这是个老字号的门派了，前几年玩 Linux 的人，几乎都用这套系统。它可完全手工打造个人需求的特性，让很多目前已是高手级的玩家仍念念不忘。Slackware 在国内用得很多，也许用来做服务器，性能会好些。最新版本的安装过程已改善了不少，想要完全掌控情况的各位朋友可以一试。

（3）Debian（http://www.debian.org）

您通常会在 Debian 字眼后看见 GNU Linux 的字样，该派别目前被大家公认为结构最严谨、组织发展最整齐，它也有一个包裹管理系统称为 DPK（Debian Package），所做的事情和 Red Hat 的 RPM 异曲同工，使整体文件的管理更加方便。Debian 的原始程序代码都是遵循 GNU 的方式开放的，所以它完全符合开放源代码精神，不像其他的 Linux 都或多或少地保留了一部分程序代码不开放（Red Hat 是直到 6.0 版才全部开放的）。

（4）SuSE（http://www.suse.com）

这是一套在欧洲相当受欢迎的版本，它和 XFree86 合作开发 x86 上的 X Server。安装 SuSE 时可以选择显示德文或英文，它还有自己的一套设定程序叫做「SaX」，可以让使用者较方便地设定。它的安装套件也采用 RPM 模式，所以要安装、升级或移除程序都非常方便，目前版本为 8.0。

（5）OpenLinux（http://www.caldera.com）

这是由 Caldera 公司推出的版本，并不是很「Open」的一个版本，网上可以下载其 Lite 版本，但正式版本是要 money 的，因为整个套件中有许多商用软件，所以并没有提供网络下载的服务。

（6）TurboLinux（http://www.pacific.com）

由 Pacific HiTech 公司开发的套件，该套件在日本市场上占有一席之地，从安装到使用接口都是日文的。在国内它与清华大学及研究机构合作研发了中文版本，在国内掀起了一股 Linux 潮流。

（7）Red Flag（http://www.redflag-linux.com）

这是由中科红旗软件技术有限公司推出的中文版本的 Linux，该 Linux 在众多的中国 Linux 用户中占有一定的比例。可以从网络上下载其红旗桌面版。

（8）Mandrake Linux

它的吉祥物是一个黑色的魔术帽，它其实是在 Red Hat 的基础上制成的，它继承了许多 Red Hat 的优点，还加上了许多迎合 Linux 初学者的功能，如美丽的图形化安装界面。7.0 版本开始走向成熟，赢得了不少用户。而且自从推出 8.0、8.1 及 8.2 版后，使得 Mandarke 也获得了较高的用户占有率。刚从 Windows 中走出来的朋友，可以一试，挺不错的。

（9）BluePoint Linux

很多人觉得这是做得最成功的一款中文 Linux 发行版，蓝点还是挺有创新、挖得挺深的一个 Linux 厂商。但是其稳定性不是太好，适于桌面，不适于做服务器。

其他还有很多发行版本，在此不做逐一介绍了。

1.2　Linux 内核版本信息

本节对 Linux 非同寻常的版本编号及 GPL 术语相关问题加以说明。

1.2.1　Linux 内核版本

首先要注意，Linux 系统使用的每个软件包都有它自己的发行号，它们之间有一些内部的依赖关系：需要在特定版本的软件包上使用特定版本的另一个软件包。Linux 发布版的开发者通常要处理大量软件包的匹配问题，而使用者从预先打包的发布版本上安装软件，无需考虑版本问题。另外，那些更换和升级系统软件的人都要自己处理版本问题。所幸的是，一些最新发布的版本允许单个软件包的升级，它是通过检查软件包间的依赖性实现的，这大大简化了使用者为维护系统软件一致所要做的事情。

注意，最近的内核包含一个称为 Documentation/Changes 的文件，它罗列了所有编译和运行这个内核版本所需要的软件。1.2 版的源码中没有这个文件。

偶数内核版本（如 1.2.x 和 2.0.x）是稳定版本，专门用于发布的版本。而相反，奇数版本是开发中的一个快照，相当短暂；最新的版本代表最新的发展状况，但可能过几天就过时了。不过有时你会选择运行一个开发用内核，或者由于它有一些在稳定版本中没有的特性，而你正好需要这些特性；或者很简单，你就是自己改动了这些版本的一些特性，并且你没有时间更新补丁。不过还是要注意，实验用内核没有什么保证，如果由于在非当前奇数版本内核中的臭虫导致你丢失数据，没人帮得了你。

本书介绍的一些特征在旧版本中是没有的。大多数样例模块在很多内核版本中都是可以编译和运行的。

由于 Linux 已不再只是“PC 兼容机的 UNIX 变体”了，它的另一个特点是，它是平台独立的操作系统。事实上，除了 x86 以外，它也成功地用于 Axp-Alpha、Sparc 处理器、Mips Rx000 和其他一些平台。本书也尽力达到平台无关性，而且所有的例子代码都在 PC、Alpha 平台和 Sparc 机器上测试过。由于代码在 32 位和 64 位处理机（Alpha）上测试过，它们应该可以在其他平台上编译和运行。

1.2.2　GPL 相关术语

Linux 按 GNU GPL（General Public License）分布许可证，这是由自由软件基金会为 GNU 计划设计的文档。GPL 允许任何人重新发布、出售 GPL 的产品，只要允许接收者从源码重新建立二进制文件的精确副本。另外，任何从 GPL 产生的软件产品也必须按 GPL 发布。

这种许可的主要目的是通过允许每个人随意修改程序来推广知识；同时，向公众出售软件的人仍可以做他们的工作。尽管这是个很简单的目标，但还是有一些正在进行的有关 GPL 及其使用的讨论。如果想读到这些许可证，你可以在自己系统的若干个地方找到它们，包括目录/usr/src/linux，有一个称为 COPING 的文件。

当涉及第三方和定制的模块时，它们不属于内核，因此你无须限制它们使用 GPL 许可证。模块通过明确的接口使用内核，它不是内核的一部分，这种关系和用户程序通过系统调用使用内核很相似。

简而言之，如果你的代码深入内核内部而你又想发布代码，则必须使用 GPL。尽管自己修改、自己使用不是非要使用 GPL，但如果你要发布代码，就必须在发布中包含源代码——人们获得你的软件包，并且可以随意修改它，重建二进制文件。换句话说，如果你写了一个模块，你就可以按照二进制格式发布此模块。然而，模块通常针对要连接的内核重新编译，这也不总是可行的。对于发布模块的二进制代码，一般障碍是：模块包含了定义或声明在内核头文件中的代码；不过这个障碍并不能成立，因为头文件是内核公共接口的一部分，因此它不受许可证制约。

当例子中包含了部分内核代码时，GPL 就适用了：与源码一同发行的文本很清楚地说明了这一点。这仅对某些源文件是有效的，对本书主题而言，这是次要的。

1.3　Linux 内核体系结构

Linux 内核主要由 5 个模块构成，分别是：进程调度模块、内存管理模块、虚拟文件系统模块、进程间通信模块和网络接口模块。

图 1.1 说明了 Linux 内核的重要组成部分，以及它们之间的关系。

下面简单介绍一下各个模块的作用。

（1）进程调度模块

进程调度程序是内核的重要组成部分，它选择下一个要运行的进程并负责控制进程对 CPU 资源的使用。调度程序采用一种策略使各个进程能够公平合理地访问 CPU，同时保证内核能够实时地执行必要的硬件操作。

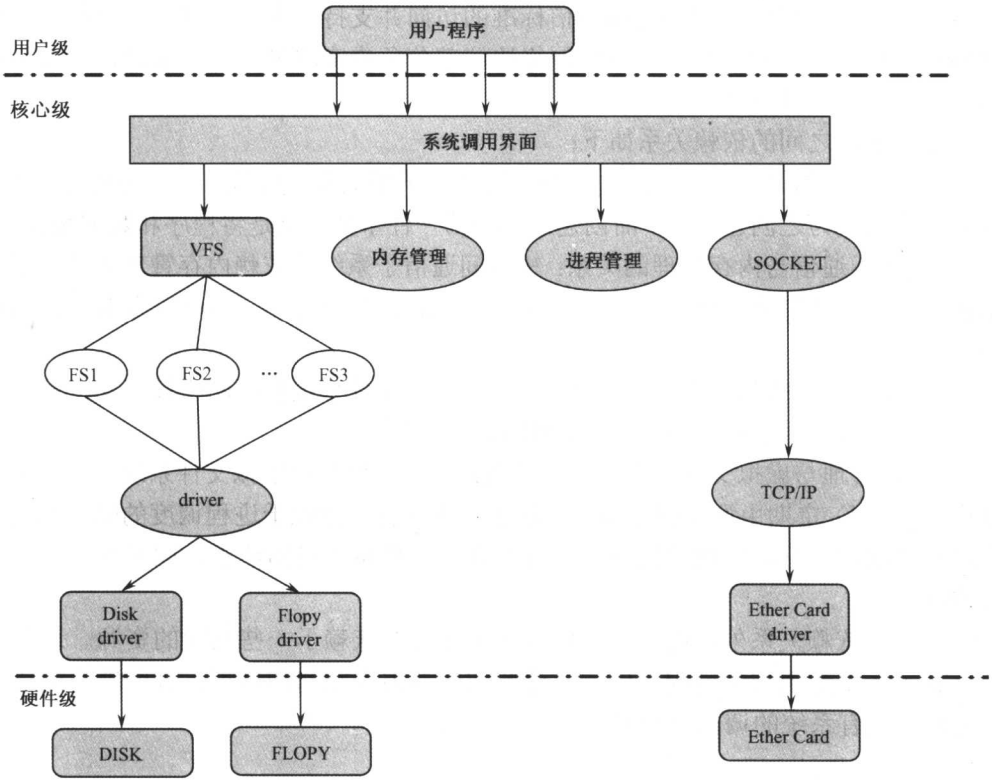

图 1.1　Linux 内核体系结构

（2）内存管理模块

内存管理模块负责管理系统，用于确保所有进程能够安全地共享计算机的内存。同时，内存管理模块还支持虚拟内存，使得 Linux 能够支持进程使用比实际内存空间更大的内存地址空间。

（3）虚拟文件系统模块

虚拟文件系统（VFS）模块通过向所有的外部存储设备提供一个通用的文件接口，隐藏了各种硬件的不同细节，从而提供并支持与其他操作系统兼容的多种文件系统。Linux 最好的特征之一就是它支持多种文件系统，用户不仅可以从自己的文件系统如 ext2，ext3，ReiserFS 等查看文件，而且可以从与其他操作系统相关的文件系统查看文件。对用户来说，从一种文件系统到另一种文件系统没有任何差异，只要是 Linux 支持的文件系统类型，用户就可以很方便地将它安装到 Linux 系统中使用。

（4）进程间通信模块

进程间通信模块主要负责进程之间如何进行信息交换或共享信息等工作。Linux 提供了多种进程之间的通信机制，其中信号和管道是最基本的两种。此外，Linux 还提供了 System V 特有的进程间通信机制，包括消息队列、信号量、共享内存等。为了支持网络通信，Linux 还引入了套接字（Socket）机制。

此为试读，需要完整PDF请访问：www.ertongbook.com