深入分析 Linux 内核源代码

陈莉君　编著

人民邮电出版社
POSTS & TELECOMMUNICATIONS PRESS

深入分析 Linux 内核源代码

陈莉君　编著

人民邮电出版社

图书在版编目（CIP）数据

深入分析 Linux 内核源代码/陈莉君编著. —北京：人民邮电出版社，2002.8

ISBN 7-115-10525-1

I. 深… II. 陈… III. Linux 操作系统 IV. TP316.89

中国版本图书馆 CIP 数据核字（2002）第 056978 号

内 容 提 要

自由软件 Linux 操作系统源代码的开放，为我们掌握操作系统核心技术提供了良好的条件。本书共分 13 章，对 Linux 内核 2.4 版的源代码进行了较全面的分析，既包括对中断机制、进程调度、内存管理、进程间通信、虚拟文件系统、设备驱动程序及网络子系统的分析，也包括对 Linux 整体结构的把握、Linux 的启动过程的分析及 Linux 独具特色的模块机制的分析与应用等。其中重点剖析了 Linux 内核中最基础的部分：进程管理、内存管理及文件管理。

本书对于那些准备进入 Linux 操作系统内部，阅读 Linux 内核源代码以及在内核级进行程序开发的读者具有非常高的参考价值。同时，操作系统实现者、系统程序员、Linux 应用开发人员、嵌入式系统开发人员、系统管理员、在校的大学生和研究生及对 Linux 感兴趣的用户均可在阅读本书中受益。

深入分析 Linux 内核源代码

◆ 编　　著　陈莉君

责任编辑　魏雪萍

◆ 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@ptpress.com.cn

网址　http://www.ptpress.com.cn

读者热线　010-67180876

北京汉魂图文设计有限公司制作

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：34.25

字数：831 千字　　2002 年 8 月第 1 版

印数：1-5 000 册　　2002 年 8 月北京第 1 次印刷

ISBN7-115-10525-1/TP・3021

定价：54.00 元

前　言

如果说 Linux 的出现是一个偶然，那么，席卷全球的 Linux 热潮则是一个奇迹，Linux 正以势不可挡的趋势迅猛发展，其发展前景是无法预测的。

有人说，“Linux 不就是类 UNIX 吗？”是的，它的外在表现形式确实与 UNIX 完全兼容，这也是它赖以生存的基本条件。但是，它的内涵则完全不同，这首先体现在其源代码全部重写及开放，其次是它的快速更新和发展，而更重要的是世界范围内众多计算机爱好者能通过 Internet 参与开发，由此可见，借助于 Internet 的肥沃土壤，Linux 的迅速发展是毫无置疑的！

实际上，Linux 最本质的东西体现在其“自由”和“开放”的思想，“自由”意味着世界范围内的知识共享，而“开放”则意味着 Linux 对所有的人都敞开大门，在这开放而自由的天地里，你的创造激情可以得到充分的发挥。

Linux 内核源代码的开放给希望深入操作系统内部世界的人敞开无私的胸怀，我们有幸走进了这个世界，这是一个神奇、错综复杂而又充满诱惑的世界，让喜欢迎接挑战的人们可以充分检验自己的勇气和耐力。

Linux 内核全部源代码是一个庞大的世界，大约有 200 多万行，占 60MB 左右的空间。因此，如何在这庞大而复杂的世界中抓住主要内容，如何找到进入 Linux 内部的突破口，又如何能把 Linux 的源代码变为自己的需要，这就是本书要探讨的内容。

首先，本书的第一章领你走入 Linux 的大门，让你对 Linux 内核的结构有一个整体的了解。然后，第二章介绍了分析 Linux 源代码应具备的基本硬件知识，这是继续向 Linux 内核迈进的必备条件。中断作为操作系统中发生最频繁的一个活动，本书用一章的内容详细描述了中断在操作系统中的具体实现机制。

众所周知，操作系统中最核心的内容就是进程管理、内存管理和文件管理。本书用大量的篇幅描述了这三部分内容，尤其对最复杂的虚拟内存管理进行了详细的分析，其中对内存初始化部分的详细描述将对嵌入式系统的开发者有所帮助。

在对 Linux 内核有一定了解后，读者可能希望能够利用内核函数进行内核级程序的开发，例如开发一个设备驱动程序。Linux 的模块机制就是支持一般用户进行内核级编程。另外，读者在进行内核级编程时还可以快速查阅本书附录部分提供的 Linux 内核 API 函数。

网络也是 Linux 中最复杂的部分之一，这部分内容足可以写一本书。本书仅以面向对象的思想为核心，分别对网络部分中的四个主要对象：协议、套接字、套接字缓冲区及网络设备接口进行了分析。有了对这四个对象的分析，再结合文件系统、设备驱动程序的内容，读者就可以具体分析自己感兴趣的相关内容。

Linux 在不断地发展，本书介绍的版本为 Linux 2.4.16。尽管本书力图反映 Linux 内核较本质的东西，但由于笔者的知识有限，对有些问题的理解难免有偏差，甚至可能有“Bug”，希望读者能尽可能多地发现它，以共同对本书进行改进和完善。

在本书的编写的过程中，笔者查阅了大量的资料，也阅读了大量的源代码，但本书中反映的内容也仅仅是 Linux 的主要内容。因为一本书的组织形成是一种线性结构，而知识本身的组织结构是一种树型结构，甚至是多线索的网状结构，因此，在本书的编写过程中，笔者深感书的表现能力非常有限，一本书根本无法囊括全部。在参考书目中，我们将给出主要的参考书及主要网站的相关内容。

本书的第一版是《Linux 操作系统内核分析》，在第一版的编写过程中，康华、季进宝、陈轶飞、张波、张蕾及胡清俊等参与了编写。第一版出版后得到了很多读者的充分肯定和赞扬，并授权台湾地区出版。在本次改版的过程中，依然保留了第一版的风格，但加深了对进程管理、内存管理及文件管理等众多内容的剖析。

这次改版由于时间仓促，加之作者的水平有限，书中有些术语的表达可能不妥，有些内容的分析也可能不够准确，敬请读者朋友批评指正。我的联系方式是：cljun@xiyou.edu.cn。

编者

2002 年 7 月

目　录

第一章　走进 Linux

对经常使用计算机的人来说，时常会感到操作系统是一个神奇、神秘而又几乎无所不能的“上帝”。一打开计算机，我们首先看到的是操作系统，所有软件的运行都离不开它，它给我们带来一个个的惊喜，但有时也带来烦恼和不安。

实际上，很多人都有这样强烈的愿望，即“上帝”到底是怎样操纵这一切的？UNIX 操作系统曾经敞开 UNIX 操作系统的胸怀，让我们窥视到它的内在机制，但它毕竟属于“贵族”阶层，我们大多数人并不能使用上它。

Windows 以平民的身份来到我们中间，我们欢呼它的友好和平易近人，正因为 Windows，才使得计算机走进我们寻常百姓家，使得计算机普及成为现实。但 Windows 有时也“伤风感冒”，我们想找到原因，以便对症下药。可是，Windows 的窗户并没有打开，我们无法透过窗户看看 Windows 的内部世界到底是什么样的，这让我们困惑，尤其让喜欢追根寻源的人们感到失望。

Linux 带着一股清新的风翩翩而来，它并不成熟，也不完美，甚至自身有很多缺点，可 Internet 的龙卷风把它吹遍世界，世界各地的计算机爱好者狂热地喜欢上 Linux。Linux 不再是一个孤单的个体，而成为软件发展史上的“自由女神”，很多的计算机高手和计算机爱好者为之倾其了极大的热情。它在迅速地成长，短短几年功夫，从一个摇摇晃晃的婴儿成长为脚步稳健的少年。这一切都源于什么？那就是 Linux 的创始人 Linus Torvalds 把 Linux 适时地放入到了 GNU 公共许可证下。

1.1　GNU 与 Linux 的成长

GNU 是自由软件之父 Richard Stallman 在 1984 年组织开发的一个完全基于自由软件的软件体系，与此相应的有一份通用公共许可证（General Public License，简称 GPL）。Linux 以及与它有关的大量软件是在 GPL 的推动下开发和发布的。

自由软件之父 Stallman 像一个神态庄严的传教士一样喋喋不休地到处传播自由软件的福音，阐述他创立 GNU 的梦想：“自由的思想，而不是免费的午餐”。这位自由软件的“顶级神甫”为自己的梦想付出了大半生的努力，他不但自己创作了许多自由软件如 GCC 和 GDB，在他的倡导下，目前人们熟悉的一些软件如 BIND、Perl、Apache、TCP/IP 等都成了自由软件的经典之作。

如果说 Stallman 创立并推动了自由软件的发展，那么，Linus 毫不犹豫奉献给 GNU 的 Linux，则把自由软件的发展带入到一个全新的境界。

实际上，Linus 是一个理想主义者，但他又脚踏实地。当 Linux 的第一个“产品”版本

Linux 1.0 问世的时候，是按完全自由扩散版权进行扩散的。他要求 Linux 内核的所有源代码必须公开，而且任何人均不得从 Linux 交易中获利。他这种纯粹的自由软件的理想实际上妨碍了 Linux 的扩散和发展，因为这限制了 Linux 以磁盘拷贝或者 CD-ROM 等媒体形式发行的可能，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 决定转向 GPL 版权，这一版权除了规定自由软件的各项许可权之外，还允许用户出售自己的程序拷贝。

这一版权上的转变对 Linux 的进一步发展可谓至关重要。从此以后，便有很多家技术力量雄厚又善于市场运作的商业软件公司，加入到了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，磨光了自由软件许多不平的棱角，增加了更易于用户使用的图形用户界面和众多的软件开发工具，这极大地拓展了 Linux 的全球用户基础。

Linux 内核的功能以及它和 GPL 的结合，使许多软件开发人员相信这是有前途的项目，开始参加内核的开发工作。并将 GNU 项目的 C 库、gcc、Emacs、bash 等很快移植到 Linux 内核上来。可以说，Linux 项目一开始就和 GNU 项目紧密结合在一起，系统的许多重要组成部分直接来自 GNU 项目。Linux 操作系统的另一些重要组成部分则来自加利福尼亚大学 Berkeley 分校的 BSD UNIX 和麻省理工学院的 X Window 系统项目。这些都是经过长期考验的成果。

正是 Linux 内核与 GNU 项目、BSD UNIX 以及 MIT 的 X11 的结合，才使整个 Linux 操作系统得以很快形成，而且建立在稳固的基础上。

当 Linux 走向成熟时，一些人开始建立软件包来简化新用户安装和使用 Linux。这些软件包称为 Linux 发布或 Linux 发行版本。发行 Linux 不是某个个人或组织的事。任何人都可以将 Linux 内核和操作系统其他组成部分组合在一起进行发布。在早期众多的 Linux 发行版本中，最有影响的是 Slackware 发布。当时它是最容易安装的 Linux 发行版本，在推广 Linux 的应用中，起了很大的作用。Linux 文档项目（LDP）是围绕 Slackware 发布写成的。目前，RedHat 发行版本的安装更容易，应用软件更多，已成为最流行的 Linux 发行版本；而 Caldera 则致力于 Linux 的商业应用，它的发展速度也很快。这两个发行版本也有相应的成套资料。在中文的 Linux 发行版本方面，国内已经有众多的 Linux 厂商，如红旗 Linux，BluePoint Linux，中软 Linux 等。每种发行版本有各自的优点和弱点，但它们使用的内核和开发工具则是一致的。

1.2 Linux 的开发模式和运作机制

自由软件的出现，改变了传统的以公司为主体的封闭的软件开发模式。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。

商业 UNIX 开发过程中，整个系统的开发有严格的质量保证措施、完整的文挡、完善的源代码、全面的测试报告及相应的解决方案。开发者不能随意增加程序的特性和修改代码的关键部分，如果要修改代码，他们得将其写入错误报告中才能使其有效，并随后接收源代码控制系统的检查，如果发现修改不合适，修改也可能作废。每个开发者设计系统代码的一个或几个部分，开发者只有在程序检查过程中才能更改相应的代码。质量保证部门在内部对新的操作系统进行严格的回归测试，并报告发现的问题，开发者则有责任解决所报告的问题。质量保证部门采用复杂的统计分析系统以确保在下次发行时有百分之几的程序错误已修改。

总之，商业 UNIX 开发过程使得其代码非常复杂，公司为了保证下次操作系统的修订质量，得收集和统计分析操作系统的性能。开发商业 UNIX 是一个很大的工程，常常大到有数以百计的编程者、测试员、文档员以及系统管理员参与。

对于 Linux，你可将整个组织开发的概念、源代码控制系统、结构化的错误报告、统计分析等通通扔到一边去。

Linux 最初是由一群来自世界各地的自愿者通过 Internet 共同进行开发的。通过互联网和其他途径，任何人都有机会辅助开发和调试 Linux 的内核、链接新的软件、编写文档或帮助新用户。实际上，并没有单独的组织负责开发此系统，Linux 团体大部分通过邮递清单和 USENET 的消息组进行通信。许多协定已跳过开发过程，如果你想将自己的代码包括进“正式”内核，只需给 Linus Torvalds 发一个邮件，他就会进行测试并将其包括进内核（只要代码不使内核崩溃并且不与整个系统设计相悖，Linus 都很乐意将其包括进去）。

Linux 系统本身采用彻底开放、注重特性的方法进行设计。一般规律是大约隔几个月就发行一个 Linux 内核的新版本。当然发行周期还依赖于其他一些因素，如排除的程序故障数、用户测试预发行版的返回数以及 Linux 的工作量等。

可以说在两次发行期间，并不是每个故障都已排除，每个问题都已得到了解决。只要系统不出现很挑剔或明显的故障，就认为比较稳定，可以推出新版本。Linux 开发的动力不在于追求完美、无故障，而是要开发 UNIX 的免费实现。

如果你想把新的特性或应用软件增加到系统上，就得经过一个“初始”阶段。所谓“初始”阶段，就是一个由一些想对新代码挑出问题的用户不断进行测试的阶段。由于 Linux 团体大多在 Internet 上，“初始”软件通常安装在一个或多个 LinuxFTP 上，并且在 LinuxUSENET 消息组上张贴一张如何获取和测试其代码的消息，从而使得下载和测试“初始”软件的用户可以将结果、故障或问题等邮件告之作者。

初始代码中的问题解决后，代码就进入“第二”阶段：工作稳定但还不完全（即能够工作，但可能还不具备所有特性）。当然，它也可能进入“最后”阶段，即软件已完备并且可以使用。对于内核代码，一旦它完备，开发者就可让 Linus 将其包括进标准内核内，或者作为内核的可增加选项。

注意，这些仅是达成协定，并未形成规则。很多人对他们的软件不必发行“初始”或测试版充满信心，因此发行哪个版本是根据开发者的决定而定的。

你可能对一群自愿者居然能编写、调试出完整的 UNIX 系统惊讶不已。整个 Linux 内核通过拼凑而成，没有采用专利的源代码，大量工作都由自愿者完成，他们将 GNU 下的免费软件移植到 Linux 系统下，同时开发出库、文件系统以及通用的设备硬件驱动程序等。

实际上，Linus 率领的分布在世界各地的 Linux 内核开发队伍仍然在高速向前推进。当前