异步图书
www.epubit.com.cn

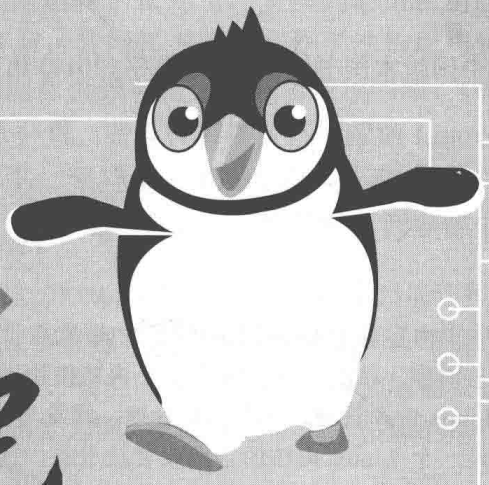
张天飞◎著

奔跑吧 Linux内核

基于 Linux 4.x 内核源代码问题分析

- 基于 Linux 4.x 内核和 Android 7.x 内核
- 基于 ARM32/ARM64 体系架构
- 以实际问题为导向的内核分析书籍
- 反映内核社区新技术发展
- 新黑科技：EAS 调度器、MCS 锁、QSpinlock、Dirty COW

中国工信出版集团
人民邮电出版社
POSTS & TELECOM PRESS



张天飞◎著

奔跑吧 Linux内核

基于 Linux 4.x 内核源代码问题分析

人民邮电出版社
北京

图书在版编目（CIP）数据

奔跑吧Linux内核 : 基于Linux 4.x内核源代码问题分析 / 张天飞著. -- 北京 : 人民邮电出版社, 2017.9
ISBN 978-7-115-46502-3

Ⅰ. ①奔… Ⅱ. ①张… Ⅲ. ①Linux操作系统 Ⅳ. ①TP316.85

中国版本图书馆CIP数据核字(2017)第162619号

内 容 提 要

本书内容基于 Linux 4.x 内核，主要选取了 Linux 内核中比较基本和常用的内存管理、进程管理、并发与同步，以及中断管理这 4 个内核模块进行讲述。全书共分为 6 章，依次介绍了 ARM 体系结构、Linux 内存管理、进程调度管理、并发与同步、中断管理、内核调试技巧等内容。本书的每节内容都是一个 Linux 内核的话题或者技术点，读者可以根据每小节前的问题进行思考，进而围绕问题进行内核源代码的分析。

本书内容丰富，讲解清晰透彻，不仅适合有一定 Linux 相关基础的人员，包括从事与 Linux 相关的开发人员、操作系统的研究人员、嵌入式开发人员及 Android 底层开发人员等学习和使用，而且适合作为对 Linux 感兴趣的程序员的学习用书，也可以作为大专院校相关专业师生的学习用书和培训学校的教材。

◆ 著　　张天飞
责任编辑　张　涛
执行编辑　张　爽
责任印制　焦志炜

◆ 人民邮电出版社出版发行　　北京市丰台区成寿寺路 11 号
邮编　100164　　电子邮件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
大厂聚鑫印刷有限责任公司印刷

◆ 开本：787×1092　1/16
印张：47.5
字数：1126 千字　　2017 年 9 月第 1 版
印数：1 – 4 000 册　　2017 年 9 月河北第 1 次印刷

定价：158.00 元

读者服务热线：(010)81055410　印装质量热线：(010)81055316
反盗版热线：(010)81055315
广告经营许可证：京东工商广登字 20170147 号

对本书的赞誉

在参加 2017 年北京举办的 LinuxCon 大会期间遇到了张天飞，了解到他正在写作一本《奔跑吧 Linux 内核》新书。回来后读了本书的样章，其问答方式的写作手法构思巧妙；以工程实践经验为基础，让读者把知识活学活用的创意也颇有特色。书名也很吸睛，《奔跑吧 Linux 内核》这个书名，源于作者每天坚持奔跑 5 公里，而且该书作者打算跟随 Linux 内核版本的演变不断地更新本书。也希望读者跟随本书，坚持学习 Linux 内核不动摇。

——陈莉君　西安邮电大学

Linux 是一个应用非常广泛的、成熟的操作系统。Linux 内核是整个 Linux 的基础和核心，包括从存储管理、CPU 和进程管理、文件系统、设备管理和驱动、网络通信到系统引导、系统调用等内容，非常值得搞嵌入式、物联网、机器人、智能硬件、VR/AR 等领域需要软硬件协同开发设计的工程师们深入研究。此书就是以 Linux 为例，详尽阐述了原本枯燥的操作系统的方方面面的知识，是一本很好的从知晓到熟悉 Linux 的进阶学习读物。

张天飞是 12 年前和我在上海亿道的同事，非常热爱底层技术探究。直到现在还能够静下心来做些底层研究的同志不多，希望他可以不断分享多年学习心得和从业经验给广大 Linux 学习者。加油！

——石庆　亿道控股 Emdoor 联合创始人&亿境虚拟现实技术有限公司总经理

Linux 内核与我们的生活息息相关，从手机、平板电脑、服务器、汽车到智能家电，都能看到它的身影。长久以来，一直没有一部深入浅出介绍整个 Linux 内核的中文书。英文书很多也是稍显过时，因为内核的变化是如此之快。很高兴看到有这样的一本书出版，把最新的内核与内核设计及一些重要变更的原因呈现出来，让内核不再是一个黑盒子。这对任何要做性能优化、开发驱动程序，甚至直接修改内核的人来说是一大福音。

——Tim Chen　Linux 内核资深技术专家

这是一本深入讲解基于 ARM Cortex-A 处理器在服务器和智能设备上运行 Linux 系统的书，可以帮助读者理解硬件如何与底层 Linux 内核交互，对 Linux 内核爱好者和 Platform/BSP 软件开发者系统学习工作很有益。

——修志龙　ARM 公司应用工程师经理

对于安卓智能手机底层系统研发人员来说，本书有如一场及时雨，不仅在全球范围内首次解读了最新的 ARM64 体系架构和 Linux 4.x 内核，还及时呈现了与智能手机系统用户体验密切相关的内核新技术，比如 EAS 调度器。本书作者携十余年的 Linux 内核和驱动开发经验，倾情奉献，诚意满满，推荐细细品读、慢慢揣摩！

——吴章金　魅族手机研发中心 BSP 部技术总监

本书的形式设计非常巧妙，它采用一种启发问答的形式，这样容易让读者带着问题去阅读，并可以直接用回答问题来验证阅读的效果。本书的另外一个特点是内容新，能够紧

扣内核的新变化。

——宋宝华　Linux 内核资深技术专家，技术畅销书作者

这是一本 Linux 操作系统工匠的力作，作者站在 Linux 操作系统前沿，以情景分析的方法向我们展示了最新版本内核的秘密。与所有深入讲解内核代码的书籍一样，本书同样值得读者反复推敲、仔细琢磨。如果你在阅读本书的过程中有更好的建议和意见，请告诉所有人。毕竟，开源社区是集市，而不是教堂。

——谢宝友　中国开源软件推进联盟专家委员，Linux ZTE 平台维护者

在软件定义一切的时代，作为开源世界重要基石的 Linux 变得越发重要，掌握坚实的 Linux 内核知识几乎是软、硬件工程师进阶所必须的。本书作者采用交互问答的方式，将最新 Linux 内核抽丝剥茧，依次呈现给读者，既适合初、中级开发人员系统学习，也适合高级开发人员随时参阅，强力推荐！

——段夕华　IT 老兵，开源技术爱好者

伴随计算机层次化体系结构的更迭，操作系统、编译系统和数据库作为 IT、互联网及物联网的基石，多年来不断演进。而 Linux 内核自 1991 年发起至今，集数万人智慧结晶，承上启下，早已成为学术界与工业界协作与创新的重要平台。本书作者从事 Linux 内核研发多年，勤于总结，故能将其脉络梳理详略得当，恰到好处。希望本书会让您踏上一次愉悦的内核之旅，不虚此行。

——刘杰　百度主任研发架构师，Linux 内核资深技术专家，XFS 文件系统核心开发者

学习 Linux 内核的第一手材料必然是代码，但是单纯研读代码犹如盲人摸象，容易迷失方向。本书立足于代码分析，辅以大量的子系统的概观，并以启发式问题为线索，让你在 Linux 内核的世界游刃有余、得心应手。

——赖江山　Linux 内核 SRCU 模块的维护者

大数据与人工智能的发展方兴未艾，遮掩了 TMT 底层基础设施应有的光芒。Linux 从 1991 年至今，廿年有余，历经了最初的前卫与今日的普及，每一个年代依然在演绎着新的故事。辉煌之余，略有遗憾，近些年全球鲜有书籍对 Linux 4.x 时代进行系统的梳理，本书弥补了这一遗憾，在此向致力于底层基础架构领域的读者推荐此书。

——王齐《Linux PowerPC 详解——核心篇》和《PCI Express 体系结构导读》作者

毫无疑问，ARM 平台是目前使用最广泛的计算机平台，也是 Linux 系统应用最广泛的平台，这本基于 ARM 的 Linux Kernel 4.x 内核分析来得恰是时候。本书从 ARM 的系统硬件开始介绍，导出基于这些硬件的内核软件设计；从应用常见的系统调用开始，展开到在内核中如何实现这些系统调用，为中级层次读者一一揭开 Linux 系统内核的面纱。独特的问答方式也为该书的一大亮点，即使是内核老手也能在阅读中发现乐趣。希望此书能给国内广大内核爱好者带来欢乐和帮助！

——时奎亮　Linaro 资深内核专家

推荐序一

As Linux spreads out into more and more systems in all areas of computing, understanding the internals of the operating system becomes a very valuable skill. This book will help you learn about the core internals of the Linux operating system, providing you the knowledge to be able to adapt Linux to work properly for the new devices and environments that you create.

Linux 操作系统已经部署到越来越多计算领域的系统中，理解操作系统内核的实现就变成一个具有极高价值的技能。《奔跑吧 Linux 内核》可以帮助你学习 Linux 操作系统最关键的内核，让你有足够多的知识去将 Linux 顺利应用到你所创造的新设备和新应用环境中。

—— **Greg Kroah-Hartman**

Greg Kroah-Hartman 简介：Linux 基金会院士，Linux 内核核心领袖之一，Linux stable tree 的维护者，《Linux Device Drivers》一书的作者之一。

推荐序二

非常荣幸接到张天飞的邀请，为《奔跑吧 Linux 内核》一书写序。

初识天飞，大概是十几年前了。那时的天飞大学毕业不久，我已经当了十多年的大学教师。由于共同的爱好和热情，我们有缘在计算机底层系统软件，尤其是 Linux 操作系统内核这一神秘而充满乐趣的领域中一起摸爬滚打、专研内核技术。跟他的名字一样，天飞给我的印象就像一个活力四射的雄鹰，有着渴望求知的翅膀，永远不知疲倦地在 Linux 内核这一广阔天空自由自在地翱翔。虽然我年长于天飞，但是我们习惯称呼他为“飞哥”，因为他有一个很酷的网名叫 Figo，我猜想他是足球天才菲戈的粉丝。又正巧我也非常喜爱足球，这加深了我们惺惺相惜的战斗情谊。十几年前，我们俩在一个“战壕”里工作了很长一段时间，并且合作出版了一本嵌入式系统相关的教材书籍。

转眼间，当年的飞哥如今已经成为稳健成熟的“笨叔叔”，从事 Linux 内核和驱动开发有十余年的时间，也曾在多家芯片公司从事过手机芯片底层软件开发和客户支持工作，还从事 Android 手机底层软件开发和项目管理工作。十几年的技术浸润，使得他从身体到灵魂都烙上 Linux 的印记。从一个飞天少年，到一个内功深厚的 Linux“笨企鹅”，他永远在 Linux 内核的自由世界里不停地奔跑。这一次，他还要带上他的作品，跟广大读者朋友一起分享 Linux 内核的乐趣。

言归正传，说一说《奔跑吧 Linux 内核》。在物联网、大数据、云计算这些充满创新的领域，操作系统作为计算机系统软件的基石，吸引着无数技术爱好者投身其中。社会在奔跑，技术也在奔跑，Linux 内核发展至今已经越来越复杂、越来越庞大。许多新技术、新算法、新补丁不断融入到 Linux 内核之中，同时也有许多内核初学者和开发工程师加入到研究 Linux 内核的队伍之中。要充分阅读和理解 Linux 内核代码越来越不容易。各种 Linux 内核学习经典著作如同不灭的火种，点燃学习者思想的火把，使他们在 Linux 内核这条崎岖不平的道路上勇敢追寻理想、探索光明。这些经典著作，我认为大致可以分为 3 类。

（1）内核原理类：从理论层面上为读者介绍操作系统设计与实现中所涉及的技术原理，代表作有《操作系统：精髓与设计原理》《现代操作系统》《操作系统概念》。

（2）内核剖析类：从代码实现角度为读者分析操作系统主要模块的设计与实现，代表作有《FreeBSD 操作系统设计与实现》《Linux 内核设计与实现》《深入理解 Linux 内核》。

（3）动手实践类：从零开始带领读者实现一个小型内核，代表作有《Orange's：一个操作系统的实现》《30 天自制操作系统》，以及我的拙著《操作系统设计与实现》。

与上述这些书相比，《奔跑吧 Linux 内核》有着自己的独特之处。

第一，该书采用问题导向式的内核源代码分析方式。这是非常有益的尝试，颠覆了传统内核分析书籍的做法。我们都知道，Linux 内核代码动辄几百万行，阅读起来时间成本呈指数式上升，难免会让读者望而却步或者昏昏欲睡。本书作者创新性地在每一章的开头以提问的方式抛出相应问题，以吸引读者的注意力和好奇心。而且这些问题非常有趣并且贴近读者需求，它们有的来源于作者长期实际工程项目中遇到的问题并抽象总结，有的是作者在阅读和学习内核代码时产生过的疑问，有的是作者及其朋友在相关面试中关于 Linux

内核的题目。

第二，该书基于最新的 Linux 内核版本，力求反映 Linux 内核社区最新的开发技术，一些热点话题令我印象深刻，例如内存管理漏洞 Dirty COW 的分析、手机操作系统 Android 7.1.1 中各种新算法等内容。

第三，作者别出心裁地在本书开篇提供一份 Linux 内核奔跑卷，读者可以将它作为水平测量、面试题目准备之用，希望能提高读者兴趣，让读者在快乐中开始奔跑。

第四，该书内容选择少而精，以 ARM32 和 ARM64 体系结构为基础，重点介绍了 Linux 内核中最基本最常用的内存管理、进程管理、并发与同步、中断管理等模块。

相信本书的特色和内容将使读者受益匪浅。

自由软件的精神在天上飞，Linux 的企鹅在地上跑。非常诚挚地欢迎大家跟着昔日的“飞哥”、现在的“笨叔叔”一起翱翔、一起奔跑！

“奔跑吧！Linux 内核学习者！”

陈文智

2017 年 6 月于浙江大学

推荐序三

对于徘徊在 Linux Kernel 大门外的初学者而言，这个结构复杂的庞然大物无疑令人心生敬畏，既渴望能早日如庖丁解牛般游刃有余地应用，同时也感觉学起来千头万绪、无从下手。这时，一本好的入门书籍就尤为重要，它能在古树参天、藤蔓缠绕的丛林中为你开辟出一条条穿行的道路，让你从容地游走其间，赏奇景、悟真谛。

对于我学习 Kernel 的经历而言，毛德操和胡希明老师的《Linux 内核源代码情景分析》就是这样一本好书，我一直把它奉为 Linux Kernel 学习的“圣经”。初学时，我把这本书当作代码阅读的参考书，它为理解代码提供了充足的硬件和软件知识背景，在我一筹莫展时有如长者般在耳边娓娓道来。

后来从事 Linux Kernel 开发的工作，在开源社区里摸爬滚打了很多年，也有了一些自己的积累。经常遇到年轻的初学者让我推荐学习的资料，在我内心排在首位的还是《Linux 内核源代码情景分析》，然而 Linux Kernel 日新月异，架构设计不断演进，新的特性层出不穷，基于 2.4 版本 Kernel 的源代码情景分析是否依然是初学者的最佳“导师”？我犹豫了，我抑制了内心强烈推荐的欲望，因为我不确定是否会误人子弟。

我和天飞在一个技术会议上认识，他给我的第一感觉是知识面很广，同时也很注意细节。后来有幸在同一家公司工作，交流愈发频繁起来。在他向我描绘内心的愿望时，我其实有一些震撼。他认为现在内核的学习曲线越来越陡峭，硬件平台之间的竞争也越来越激烈。他希望能总结他在学习和工作中的经验，让更多人特别是非主流平台的开发者看到不同平台上的 Linux Kernel 的风景。在现在这个浮躁的年代，很多人都追求“短、平、快”，写书是一件很耗时而且有可能费力不讨好的事情。但我知道，现在 ARM 平台基于最新 kernel 的技术书籍非常欠缺，我也期望有一本书能传承情景分析，同时弥补情景分析的不足，使更多的人受益。

后来，看着基于当前最新的 4.x Kernel 的《奔跑吧 Linux 内核》逐渐成型，我内心充满期待。它同情景分析类似，以背景总览起步，以核心代码分析为辅，穿插介绍其他相关的知识点，慢慢地展开某一个子系统的优美画卷，为刚开始阅读 Kernel 源代码的初学者带来了福音。另外在开篇时设问，让读者能带着疑问读下去，在阅读的过程中努力发掘问题的答案，最后与作者给出的答案做对比来确认自己的理解是否有偏差。

当然一本书不可能解决读者的所有问题，但一本好书能带领读者走进 Linux Kernel 世界的大门。“纸上得来终觉浅”，最好的学习 Linux Kernel 的方式还是阅读源代码，并参与到真正的工程实践中来。希望《奔跑吧 Linux 内核》作为一个很好的“引路人”，为 Kernel 代码的学习者扫清障碍，引发更深层次的思考。愿你们能够早日亲睹 Kernel 的真正面目！

肖光荣

2017 年 6 月

推荐序四

Linux 及开源软件（Open Source Software）这两个名词对于笔者及各位专家应该是十分熟悉，但对一般人而言，这两个名词仍是比较陌生的。我们经常提及手机的操作系统安卓（Android）、智能家居、车载系统等，许多产品都在应用 Linux 及开源软件。每天到微博、微信、优酷翻一翻不同的信息、新闻及视频也成为了我们生活的一部分。这些视频的主角往往在很短的时间内成为“网红”，这已经是见怪不怪的事情了。想象一下，全世界究竟有多少网民与你同一时间一起关注这个视频？又有多少大事同时发生？在这个瞬息万变的互联网时代，要处理和分析这些大数据（Big Data），都要靠 Linux 及开源软件。

我本想以“日新月异”来形容科技的发展，但现在用“分秒必争”应该更为合适！在急促发展的科技背后，有无数开源社区和贡献者的参与和支持，他们不断地推动开源的发展。开源社区是一个极为多元化的世界，在社区中，大家不谈背景、性别、出身，只要志同道合，大家便可以一起参与和协作不同的开源项目。

开源社区的力量有多大？Linux 基金会对旗下所有的开源项目进行统计，截至 2015 年 8 月 31 日，共有超过 3000 名的贡献者累积贡献了 1.249 亿行的代码，这相当于 44918 人一年的工作量。假设 1000 名开发者各自进行开发，也需要 45 年才可以完成这项创举！一般来说，大家可能每月或每周要对手机系统或 App 进行更新，如果没有这么强大的社区协作，如何可以跟得上这般急促的步伐？

谈到今年（2017 年）开源社区的活动，其中一个非常有影响力的当属首次在中国举办的 LinuxCon+ContainerCon+CloudOPEN (LC3) 会议，Linux 及 Git 的创始人 Linus Torvalds 为此次会议首次访问中国，并与世界各地的开源专家一起对 Linux 及开源的主题进行交流。

我小时候很喜欢看“哆啦 A 梦”“龙珠”这些卡通片，里面会出现如竹蜻蜓、个人宇宙飞船、AI 人工智能（Artificial Intelligence）等神奇的工具，当时听起来好像是天方夜谭，但从现今的科技来看，有一些很火的开源项目，如 IoT（物联网）用于汽车及飞机无人驾技术等，这些科技产物在不久的将来即可实现。

《奔跑吧 Linux 内核》是一本难得的讲解 Linux 内核好书，也是首本 Linux 4.x 内核书籍，反映了 Linux 内核社区的科技发展，是一本体现了全球华人参与 Linux 内核社区的杰作。本书中对 Linux 内核独特的问题导向式的批注和奔跑卷让我印象深刻，可以让读者全面了解内核的工作原理和机制，让更多的人参与到 Linux 内核开发和产品开发中。这本书将让你对开发 Linux 核心有更进一步的理解及思考，我极力推荐这本书给有志成为开发人员或对 Linux 开发感兴趣的人员阅读！

开源科技渐渐成为人类的必需品之一，在此亦非常感谢如作者般的开发人员，你们就是创新科技的“开拓者”！最后，我们也希望和鼓励更多年轻的人们加入我们，一起为创新科技和开源的生态圈做出贡献！

Maggie Cheung
Linux Foundation APAC
2017 年 5 月于香港

前　言

近些年来，使用安卓操作系统的智能手机热销，未来也将是物联网、大数据、云计算的大时代，而运行在这些相关产品最深处的几乎都是 Linux 内核。我一直在凝望你，你看不见我，我是谁？我是奔跑中的 Linux 内核小企鹅。

说起和 Linux 的渊源，要追溯到十几年前的大学时代了。2002 年，正在读大二的我购买了人生第一台电脑，AMD 的毒龙 CPU，在同学的指导下安装了 RedHat 9，这是我第一次接触 Linux 操作系统。从此，我就被 Linux 系统深深地吸引了，乐不思疲地折腾着我的 RedHat 9。2004 年春天在实验室忙着做毕业设计时，蓦然回首看到小伙伴桌上有两本厚厚的《Linux 内核源代码情景分析》，我再次被深深地吸引了，心里嘀咕着不知道什么时候才能看得完和看得懂。到了 2017 年的今天，已经毕业 12 年多了。12 年的光景让我从一名大学生变成了“笨叔叔”，也让 Linux 内核这个小企鹅变成今日的科技明星。与 12 年前相比，Linux 内核代码已经发生了翻天覆地的变化，但是不变的是一群热爱 Linux 内核的小伙伴，那是一群奔跑着的年轻人。《Linux 内核源代码情景分析》这本书在 Linux 内核圈里被称为经典，可是它讲述的内核版本是 2001 年发布的 Linux 2.4.0，距今已经有 16 年了。

回顾学习 Linux 内核的那段经历，我愈发体会到 Linux 内核的功夫在 Linux 内核之外。Linux 内核变得越来越庞大，特别是现在硬件的发展速度非常快，各种不同的思想和实现如雨后春笋一般，各种各样的补丁也让人眼花缭乱。对于一个初学者或者有经验的工程师来说，要阅读和理解最新版本的 Linux 内核变得越来越困难。而且现在市面上 Linux 内核书籍都比较旧，最经典的《深入理解 Linux 内核》讲述的是 Linux 2.6.11 内核，它发布于 2005 年，《深入 Linux 内核架构》中讲述的 Linux 2.6.24 内核是 2008 年 1 月发布的。以每 2～3 个月发布一个 Linux 内核新版本的速度，这些书中的内核版本与当前的 4.x 内核不可同日而语。另外，我发现身边不少朋友很想把 Linux 内核吃透，然后购买了不少 Linux 内核的书籍，但有时好几天也没读几页。究其原因是，现在市面上已有的 Linux 内核书籍大多是教科书式地讲述知识点，机械式地讲述内核代码的实现，读起来很容易让人犯困。

Linux 内核代码由一个一个补丁组成，这些补丁都是为了解决某个问题或者添加某些新的功能，因此最好的学习方法是：理解代码是为了解决什么问题，如何解决的，要了解问题的来龙去脉。对于学习 Linux 内核这件事情来说，应该和孩提时读“十万个为什么”一样，以问题为中心，通过阅读代码和书籍来寻找答案，比如你在用 C 语言写一个很简单的程序时，应该想想 malloc 何时分配出物理内存。当你带着疑问去阅读代码以及独立思考之后，会得到一种享受和愉悦，这就是我说的“Linux 内核的功夫在于内核之外”。因此，站在设计者的角度来提出疑问，进而阅读代码和分析推理求索之后，终于有种“拨开迷雾见天日”的喜悦。

本书特色

1．问题导向式的内核源代码分析

Linux 内核庞大而复杂，任何一本厚厚的 Linux 内核书籍都可能会让人看得昏昏欲睡。

因此本书想做一个尝试，总结我多年来在学习 Linux 内核代码和实际工程项目中遇到的比较常见的疑问，以疑问为中心讲述内核代码。在讲述每章之前，首先列举出一些思考题，激发读者探索未知的兴趣。

这些思考题主要来自于如下 3 个方面。

- 从我多年来实际工程项目中遇到的问题抽象出来。我们在实际产品的研发过程中，比如手机项目研发或者其他智能产品的研发，难免要编写驱动或者系统优化，那么常常会遇到一些问题。如果对内核了解很透彻，解决问题的速度也会明显提高。例如在书中提到的驱动代码内存越界访问的问题，如果对内存管理和内核调试很熟悉，可能用几个小时就能修复 bug 了，如果换成一个不熟悉的工程师也许耗费很长时间还是找不到方向。系统中一些问题可能会是定时炸弹，随时可能引爆，因此绝大多数情况下，查找问题花费的时间要远远多于提前静下心来搞懂 Linux 内核机制的时间。
- 我在阅读内核代码时产生过的一些疑问。
- 我和身边的朋友在参加面试时经常会被问到的有关 Linux 内核的问题。

2. 力求反映 Linux 内核社区最新的开发技术

本书基于Linux 4.x 内核，我会在每章末尾尽量把内核技术的最新发展情况分享给读者。另外，我也会加入一些最新的热点话题，比如内存管理漏洞——Dirty COW 的分析；手机领域最新 Android 7.1.1 版本中的 EAS 节能调度器、WALT 算法、PELT 算法改进、Queued Spinlock 等。

3. Linux 内核奔跑卷

本书开篇会提供一份 Linux 内核奔跑卷，这也是 Linux 内核书籍中一个新的尝试。读者可以将其用于 Linux 内核水平测试或面试题目，我希望能给读者带来阅读 Linux 内核的兴趣和探索知识的乐趣。

4. QEMU 调试环境和内核调试技巧

在阅读 Linux 内核时，大多数人都希望有一个功能全面且好用的图形化界面来单步调试内核。本书中会介绍一种图形化单步调试内核的方法，即 Eclipse+QEMU+GDB。另外，本书提供首个采用“-O0”编译和调试 Linux 内核的实验，这样可以解决调试时出现光标乱跳和<optimized out>等问题。本书也会介绍实际工程中很实用的内核调试技巧，例如 ftrace 使用、systemtap、内存检测、死锁检测、动态打印技术等，这些都可以在 QEMU+ ARM Linux 的模拟环境下做实验。

5. ARM32 和 ARM64 体系架构

本书以 ARM32 和 ARM64 体系架构为蓝本，介绍 Linux 内核的设计与实现。

本书主要内容

Linux 内核涉及的内容包罗万象，但本书不想成为一本大而全的书，因此只选取了最基本最常用的内存管理、进程管理、并发与同步和中断管理这 4 个内核模块进行讲述，力求把我所理解的东西完整记录下来。

本书中每节的内容都是一个 Linux 内核的话题或者技术点，在每节开始之前会先提出若干个问题，读者可以根据这些问题先思考，然后围绕这些问题进行内核源代码的分析，最后是对相应内容的一个小结。

Linux 内核奔跑卷一共 20 道题目，每题 10 分，一共 200 分，读者可以在 2 小时内完成。

第 1 章处理器体系结构。简单介绍 ARM32 和 ARM64 结构中一些比较常见的问题，例如 cache 组织架构、cache 一致性管理、页表访问、MMU、内存屏障等与体系结构相关的内容。

第 2 章内存管理。包括物理内存初始化、内存分配、伙伴系统、slab 分配器、malloc 内存分配、mmap 系统调用、缺页中断、匿名页面的宿命、物理页面 page 结构、反向映射、页的迁移、KSM、DirtyCOW、页面回收、内存管理数据结构框架等内容。

第 3 章进程管理。包括 fork 系统调用、CFS 调度器、PELT 算法改进、SMP 负载均衡、HMP 调度器、WALT 算法、EAS 绿色节能调度器等内容。

第 4 章并发与同步。包括原子变量、spinlock、信号量、读写信号量、Mutex、RCU 等内容。

第 5 章中断管理。包括硬件中断处理、软中断、Tasklet、workqueue 等内容。

第 6 章内核调试。包括内核单步调试、ftrace 使用、systemtap 使用、内存检测、死锁检测、动态打印技术等内容。

本书罗列的内核代码均为代码片段，显示的行号也并非源代码的实际行号，只是为行文描述方便。另外，在实际代码中有大量的注释，本书为了节省篇幅而省略了大量的代码注释，建议读者对照代码来阅读。

本书在实际代码讲解时还列举了一些关键的 patch，阅读这些 patch 有助于帮助读者理解代码。建议读者下载官方 Linux 的 git tree。下载代码命令如下：

```
#git clone https://git.kernel.org/pub/scm/linux/kernel/git/torvalds/linux.git
#git reset v4.0 -hard
```

列举的 patch 格式如下：

Linux2.6.29, commit bf3f3bc5e, <mm: don't mark_page_accessed in fault path>, by Nick Piggin.

表示在 Linux 2.6.29 中加入了此 patch，git commit 的前几位 ID 号是“bf3f3bc5e”，读者可以通过“git show bf3f3bc5e”命令来查看该 patch，该 patch 的标题是“<mm: don't mark_page_accessed in fault path>”，作者是 Nick Piggin。

由于作者知识水平有限，书中难免存在纰漏和理解错误之处，敬请各位读者朋友批评指正。我的邮箱：runninglinuxkernel@126.com。新浪微博：@奔跑吧 Linux 内核。大家也可以扫描下方的二维码，到我的微信公众号中提问和交流。

关于作者

本书作者从事 Linux 内核和驱动开发十余年，是 Linux 内核的爱好者，曾在多家芯片公司从事过手机芯片底层软件开发和客户支持工作。

写一本Linux内核方面的书籍是笔者多年来的一个小心愿。本书取名为《奔跑吧Linux内核》，一是Linux内核在蓬勃发展，全球众多杰出的公司和开发者都在为Linux内核社区开发令人激动的新功能；二是我们要不断地向前狂奔才能赶上Linux内核发展的步伐；三是作者有一个人生目标，希望每天能坚持奔跑5千米，直到80岁，因此作者希望和广大读者共勉。在开始撰写本书时Linux内核版本才到4.0，完稿时Linux内核已经发展到4.10版本了，作者选择一个整数版本Linux 4.0作为本书学习和分析的版本。谭校长有一首歌叫《八十岁后》，歌词是“总相信，八十岁后，仍然能分享好戏”，他坚持要开演唱会到八十岁。我也有一个小小的目标，希望日后Linux内核发展到5.0、6.0……版本时，《奔跑吧Linux内核》依然能以最快的速度修订和大家见面。

致谢

几经放弃，几经坚持，听着Beyond的歌，咬着冷冷的牙，坚持着心中那个万里奔跑的信念。在繁忙工作之余写书是很枯燥的，但是期间我得到了众多Linux内核社区朋友的热心帮助和鼓励。特别是陈绪、吴峰光、Tim Chen、肖光荣、李泽帆、Waiman Long、冯博群、谢宝友、杜雨阳、郭健、孟卫国、修志龙、朱辉、宋宝华、吴章金、王齐、薛坤、刘勃、刘杰、郭哲佑、郭雄飞以及胡振波，他们为本书审阅了全部或者部分稿件，提出了很多很好的意见和建议。另外还要感谢石庆、Fane Li、Law Hock Yin、周祥、何章龙、杜秉权、杨永刚、张思超、段夕华、周琰玉、涂小兵、杨晨星、杨冬冬、孟皓、王建强、王智、宋吉科、何刘宇、Tian Jun等给予的帮助。感谢南京大学软件学院的夏耐老师在KSM项目中的指导，让我对Linux内核内存管理有了更深刻的理解和认知，还要感谢任德志老师的鼓励和帮助。

本书在编写过程中得到了Linux内核社区众多杰出华人开发者以及maintainer（维护者）的鼓励和帮助，他们仔细审阅本书并提出了独特的见解和建议，这些都是无价的。在此再一次表达我的感激之情，他们都是Linux内核社区最杰出的华人代表（排名不分先后）。

陈绪　中国开源软件推进联盟常务副秘书长。

吴峰光　Linux内核社区资深技术专家，0-day内核测试项目的发起人，其预读算法和回写算法享誉内核社区。

肖光荣　Linux内核社区资深技术专家，KVM社区和Qemu社区的核心开发者和maintainer。

Tim Chen　Linux内核社区资深技术专家，Linux MCS锁和Mutex自旋等待机制的作者。

Waiman Long　RedHat公司Linux内核资深技术专家，Linux内核著名的锁专家，为读写信号量、Mutex和Queue Spinlock等做出杰出的贡献。

李泽帆　华为资深 Linux 内核技术专家，Cgroup 及 cpuset 模块的 co-maintainer。

谢宝友　Linux 内核社区 ZTE 平台维护者，中国开源软件推进联盟专家委员，《深入理解并行编程》译者。

郭健　Linux 内核资深技术专家，技术网站蜗窝科技创始人。

冯博群　Linux 内核社区资深技术专家。

孟卫国　Linux 内核资深技术专家。

杜雨阳　Linux 内核社区 CFS 调度器专家，为 CFS 中的 PELT 算法做出重大优化和贡献。

修志龙　ARM 公司应用工程师经理，精通 Cortex 系列处理器架构。

王齐　计算机体系结构资深技术专家，著有《Linux PowerPC 详解——核心篇》和《PCI Express 体系结构导读》。

宋宝华　Linux 内核资深技术专家，ARM Linux 社区 maintainer，著有《Linux 设备驱动开发详解》。

吴章金　魅族手机研发中心 BSP 部技术总监。

刘杰　百度主任研发架构师，Linux 内核资深技术专家，XFS 文件系统核心开发者。

朱辉　小米科技 Linux 内核技术资深专家，KGTP 项目发起人，GDB 项目的 maintainer。

夏耐　南京大学计算机博士，操作系统资深专家。

感谢我的领导 Liu Song 先生和 Luebbers Enno 先生对我的支持和帮助。同时感谢人民邮电出版社的张涛和张爽两位编辑的辛勤付出，才让本书顺利出版。最后感谢我的家人对我的支持和鼓励，虽然周末时间都在忙于写作本书，但是他们总是给我无限的温暖。

致敬

浙江大学计算机学院是全球最早开始从事 Linux 内核研究和教学的高校之一，非常感谢陈文智院长为本书作序，陈老师一直持续关注和鼓励本书的编写和出版，给了我很多指导性的意见和建议。

毛德操和胡希明老师编写的《Linux 内核源代码情景分析》一书是中国 Linux 内核发展史上一个永恒的经典，在此向这两位老学者和前辈致敬。此时此刻，脑海里响起了一首歌：“一追再追，只想追赶生命里，一分一秒”。愿和大家一起奔跑、一起追赶、不浪费生命的一分一秒。

张天飞

2017 年夏于上海

目　录