奔跑吧 – Linux 内核

基于 Linux4.x 内核源代码分析

- 全球首本 Linux 4.x 内核书籍
- 全球首份 Linux 内核奔跑卷
- 基于最新 Linux 4.x 内核和 Android 7.x 内核
- 基于 ARM32/ARM64 技术
- 反映内核社区最新技术发展

致敬经典

最近几年 Android 手机的热销以及未来是物联网、大数据、云计算大时代，而运行在这些产品最深处的几乎都是 Linux 内核。我一直在凝望您，您看不见我，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。在 2002 年的时候，正在读大二的我购买了人生第一台电脑，AMD 的毒龙 CPU，在同学的引导下安装了 Redhat 9，这是我第一次接触 Linux 操作系统。为此就被 Linux 系统深深的吸引了，乐不思疲地折腾着我的 RedHat 9。在 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厚厚的两本《Linux 内核源代码情景分析》被深深的吸引了，心里嘀咕不知道什么时候才能看得懂和看得完。

到了 2016 年的今天，毕业过去了整整 12 年了，虽然奔波于深圳和上海两个城市，但依然像宝贝一样珍藏着，时时拿出来翻一翻。12 年可以让人成熟成长，也让 Linux 内核这个小企鹅变成今日的科技明星。Linux 内核代码和 12 年前的已经发生了翻天覆地的变化，但是不变的一群热爱的 Linux 内核的小伙伴，那是一群奔跑着的年轻人。该书在 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，距今已经快 16 年了。

回顾这段学习 Linux 内核经历，愈发体会到 Linux 内核的功夫在 Linux 内核之外。现在 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，各种不同的实现和思想也是雨后春笋，各种各样的补丁和膏药也是让人眼花缭乱。对于一个初学者或者有不少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 内核书籍都比较老，最经典的《深入理解 Linux 内核》也只是讲述 Linux 2.6.11 内核，它发布于 2005 年距今也有 11 年，德国人巨作《深入 Linux 内核架构》里讲述的 Linux 2.6.24 内核也是 2008 年 1 月发布的，距离今天也有 8 年多，以 Linux 内核每 2~3 个月发布一个新版本的速度，它们距离当今 4.x 的内核有着翻天覆地的变化。另外我发现身边不少朋友心里很想把 Linux 内核吃透，然后购买了不少 Linux 内核的书籍，捧着厚厚的书，心里嘀咕什么时候可以读完，然而有时好几天也没读几页。现在市面的 Linux 内核书籍基本上都是教科书式的讲述知识点，机械式的讲述内核代码的实现，读起来很容易让人犯困。

Linux 内核代码都是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，因此我们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，应该和小孩喜欢打破砂锅问到底一样，以问题为中心去阅读代码和阅读书籍来寻找答案，比如你在用 C 语言写一个很简单的程序的时候，你应该想想 malloc 什么时候分配出物理内存，当你怀揣着疑问去阅读代码和独立思考之后，你得到的是一种享受一种愉悦。这就是我说的 Linux 内核的功夫在与内核之外。因此，站在设计者的角度来提出疑问，进而自己阅读代码和分析推理求索之后终于有种扒开迷雾见天日的喜悦。另外一方面，对于初学者如何提出问题发现问题以及解决问题也是比较困难的。

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 内核书籍都会让看得昏昏欲睡，因此本书想做一个尝试，总结这 12 年来在学习 Linux 内核代码以及实际工程项目中比较常见的疑问，以疑问为中心进行内核代码的讲述，因此在每章讲述之前首先把思考题先列举出来，希望能激发读者探索未知的兴趣。带着疑问和思考也许是阅读 Linux 内核代码最高效的方式了。

这些思考题主要是源自于如下三个方面：

- 从笔者这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，总是免不了要编写驱动或者系统优化，那么常常会遇到这样的情况。如果对内核了解很透彻的话，解决问题的速度也会成线性的提高。比如在书中提到的驱动代码内存越界访问的问题，如果对内存管理以及内核调试很熟悉的话，可能半天就能修复 bug 了，如果换成一个不熟悉的工程师也许也耗费很长时间也找不到方向。系统中一些问题有可能会是一个定时炸弹，随时可能引爆，尤其是过了很长时间才突然爆发出来的，因此绝大多数情况下，查找问题花费的时间要远远大于提前静下心来搞懂 Linux 内核机制要多得多，而且大多数情况下是无头绪乱点鸳鸯式地解决问题。
- 笔者在阅读内核代码时产生过的一些疑问。
- 笔者和身边的朋友在参加面试时经常会被问到 Linux 内核相关的问题。

2. 力求反映 Linux 内核社区最新的开发技术

本书基于 Linux4.0 内核，但是在每章结束时尽量把最新内核技术的发展情况奉献给读者。另外也加入最新的一些热点和话题，比如今年最热最恐怖的内存管理漏洞 –dirtycow 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 等等内容。

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。读者可以作为 Linux 内核水平测量或者作为面试题目准备也可以作为一个乐子吧，但笔者还是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

阅读 Linux 内核时大多数人都希望有一个功能全面好用的图形化界面来单步调试内核，本书介绍一种图形化单步调试内核的方法，即 Eclipse+QEMU+ARM Linux。另外本书介绍实际工程中很实用的内核调试技巧，比如 ftrace 使用、systemtap、内存检测、死锁检测、动态打印技术等等，所有这些都可以在 QEMU+ ARM Linux 的模拟环境下做实验。

5. ARM32 和 ARM64 体系结构

本书以 ARM32 和 ARM64 体系结构为蓝本来介绍 Linux 内核的设计与实现。但是最新内核有一些特性是 ARM 结构中不支持的，比如 NUMA 调度器，这个只能在 X86_64 体系结构中支持了，还好 QEMU 模拟器可以支持 X86_64 体系结构。类似的新特性还有 Queued Spinlock。

本书内容概要

本书的组织架构大致是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进行思考。然后就是围绕这些问题进行内核源代码的分析，最后是笔者对这些内容的一个总结。

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 个小时内完成。

第一章 ARM 体系结构。简单介绍 ARM32 和 ARM64 结构中一些比较常见的问题，比如 cache 组织架构、cache 一致性管理、页表访问、MMU、内存独占访问、内存屏障等和体系结构相关的内容。

第二章 Linux 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。

第三章进程调度管理。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。

第四章锁与同步。包括原子变量、spinlock、信号量、读写信号量、Mutex、RCU 等内容。

第五章中断管理。包括硬件中断处理、软中断、Tasklet、workqueue 等内容。

第六章内核调试技巧。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。

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。另外实际代码中有大量的注释，为了节省篇幅省略了大量的代码注释。建议读者可以对照代码来阅读。

感谢

写一本 Linux 内核方面的书籍是笔者多年一个小心愿。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；二是我们也要不断地向前狂奔才能赶上 Linux 内核发展的步伐；三是笔者有一个人生目标，希望每天能坚持奔跑 5 公里一直到 80 岁；因此笔者希望和广大读者一起共勉。在撰写本书时 Linux 内核才 4.0，等到完稿时 Linux 内核已经发展到 4.9 版本了，但是笔者选择一个整数版本 Linux 4.0 作为本书的学习和分析的版本。谭校长有一首歌叫《八十岁后》，歌词是这样写的“总相信 八十岁后 仍然能分享好戏”，他依然坚持要开演唱会到八十岁，那我也有一个小小的目标，希望日后 Linux 内核发展到 5.0、6.0...，《奔跑吧 Linux 内核》依然能以最快的速度修订和大家见面。

几经放弃几经坚持，听着 Beyond 的歌，咬着冷冷的牙，念着心中那个万里奔跑。由于笔者是 Linux 内核社区的吃瓜群众和段子手，本书若能博君一笑或是能解答大家在 Linux 学习和工作中小小的疑惑，已是心存感激！由于笔者知识水平有限，书中有纰漏和理解错误之处在所难免，敬请各位读者朋友批评指正。我的联系方式：figo1802@126.com，微信公众号：奔跑吧 Linux 内核。

毛德操和胡希明老师著的《Linux 内核源代码情景分析》是中国 Linux 内核发展史上一个永恒的经典，本书为这两位老学者致敬。此时此刻脑海里响起了哥哥的一首歌，“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、不浪费生命一分一秒。

张天飞

2017 年春于上海

Contents