

Yet another OS written in Rust

rCore v0.2.0 实现历程与进展

陈嘉杰 (王润基) (贾越凯)

计算机科学与技术系

Table of Contents

- ▶ rCore 是个啥
- ▶ 做了什么
- ▶ 还缺什么

Beamer 模板由 @scateu (Kang Wang) 友情提供

自我介绍

- ▶ 陈嘉杰，计 72
- ▶ Email: noc@jiegec.ac.cn
- ▶ noc = NaOCan Network Operating Center
- ▶ ac = Alternating Current Academic
- ▶ 感谢于同学的域名
- ▶ EOF

希望大家

- ▶ 随时提问，因为很多同学没有学过 OS，有很多额外的东西可以讲
- ▶ 随时质疑，因为 rCore 真的很粗暴，很多东西要么没实现要么偷工减料
- ▶ 帮忙填坑，因为坑越填越多，越写后面的越发现前面写得有问题

背景故事

- ▶ 2018 年 1 月 15 日 Stanford CS140e 开始
- ▶ 第一个 Rust 的 OS 课程
- ▶ 寒假在家闲着无聊，买了 Raspberry Pi 3
- ▶ 开始跟着 CS140e 学习 Rust 和 OS
- ▶ 写了系列博客《近来做 Stanford CS140e 的一些进展和思考》

背景故事 (herox)

- ▶ 一年后的 2019 年 1 月期末考前，听说 OS 课在准备把 rCore 改造成 uCore 的代替物
- ▶ 主动请缨找路橙助教请求参与
- ▶ 又是一个闲着的寒假，开始提交最初的代码，

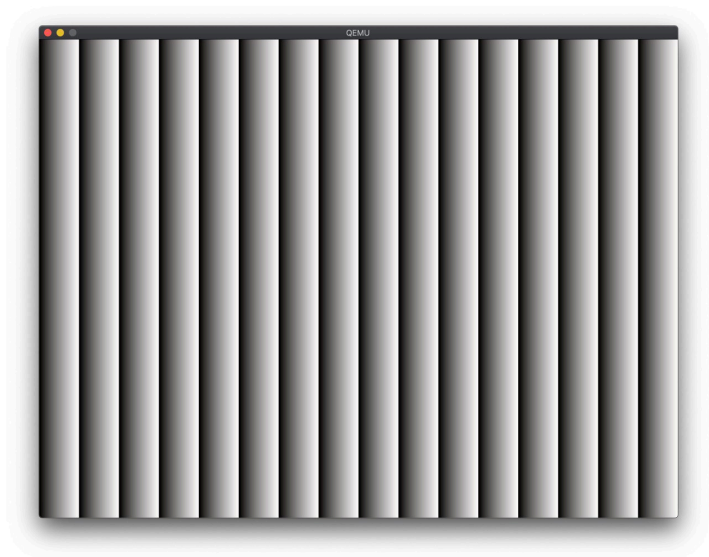
环境：RISC-V32 QEMU

- ▶ Backtrace i.e. ucore lab1
- ▶ VirtIO Net 网卡驱动，手写了一个 ping 和一个 udp 服务器作为 PoC
- ▶ VirtIO Gpu 显卡驱动，显示了一个 mandelbrot 图像
- ▶ 采坑：物理地址和虚拟地址的映射、对文档的理解

第一次能 ping 通 rCore

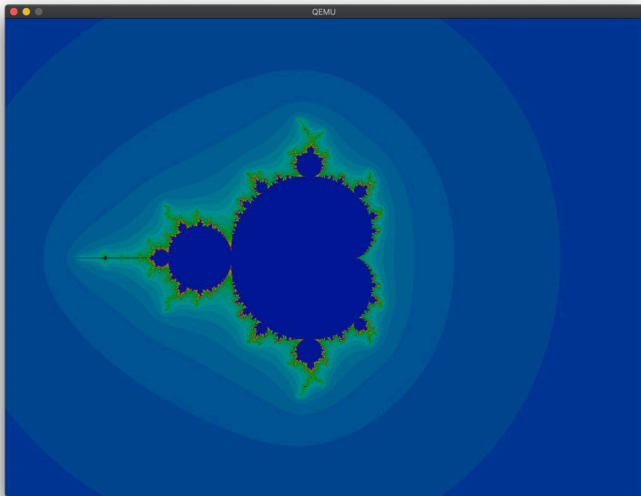
```
macbookpro ... > Data > rcore_plus > kernel > ping 10.0.0.2
PING 10.0.0.2 (10.0.0.2): 56 data bytes
64 bytes from 10.0.0.2: icmp_seq=0 ttl=64 time=4.041 ms
64 bytes from 10.0.0.2: icmp_seq=1 ttl=64 time=3.916 ms
```


第一次可以显示内容



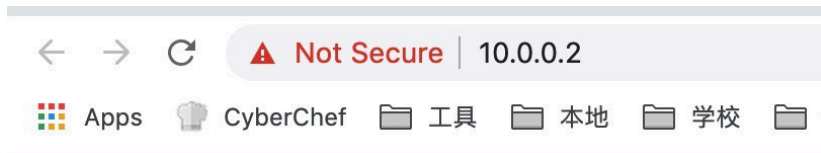
第一次显示 mandelbrot

感谢 @shankerwangmiao 的颜色建议



接入 TCP/IP 网络协议栈

- ▶ 纯 Rust 网络协议栈 smoltcp
- ▶ 实现了 udp 和 tcp 服务器



Hello, world!

实现 VirtIO Blk 驱动

- ▶ 原本的文件系统是直接二进制存放 ELF 中，
变成读 VirtIO Blk 设备
- ▶ 踩坑：对文档理解还是不准确

放假结束后回到学校

- ▶ 咕
- ▶ 咕
- ▶ 咕
- ▶ 咕
- ▶ 咕
- ▶ 不能咕了，起来干活！

出现了新的目标

- ▶ MIT CSAIL Biscuit OS
 - ▶ Go 写的操作系统，支持多核，文件系统，CoW，TCP/IP, AHCI SATA, IXGBE Driver
 - ▶ 可以跑 Nginx Redis
 - ▶ OSDI 18 “The benefits and costs of writing a POSIX kernel in a high-level language”
- ▶ 用 Rust 实现相似的功能，比较性能

快速开发期

- ▶ 一周 40+ 人时的开发时间
- ▶ codebase 翻倍
- ▶ 实现了大量的新功能
- ▶ 填了很多坑
- ▶ 成为后面的坑的 codebase

6 Weeks Ago

- ▶ 移植网卡驱动到 x86_64
- ▶ 为用户程序实现 `stdin/stdout/stderr`
- ▶ 一开始思路：同时支持 Biscuit 的 `litc` 和 `musl` 以支持 Linux 程序
- ▶ 问题：编号、参数不一致
- ▶ 一致采用 Linux 的 `syscall` 标准，用 `musl` 编译 `biscuit` 提供的部分用户态程序
- ▶ 非常正确的决定

5 Weeks Ago

- ▶ 实现 `sys_arch_prctl` 以支持 TLS on x86_64
- ▶ 开始可以运行 `musl` 编译的简单程序
- ▶ 完善文件相关一系列 `syscall`: `sys_stat`
`sys_readv` `sys_writev` `sys_lseek`
- ▶ 实现 PCI 总线设备的枚举、BAR 的获取和中断的的设定 ref: BiscuitOS
- ▶ 实现 `e1000` 网卡驱动 ref: JudgeDuck

4 Weeks Ago

- ▶ 实现网络相关 syscall: `sys_socket`
`sys_close` `sys_connect` `sys_write`
`sys_read` `sys_sendto` `sys_recvfrom`
- ▶ 完善文件系统相关 syscall: `sys_sync`,
`sys_truncate`, `sys_link`, `sys_unlink`
- ▶ 支持了 ping: raw socket; 支持了 udp 和
tcp: nc
- ▶ 添加了 RTC 驱动, 完善了时间相关的
syscall
- ▶ `date`、`ls` 可以工作了

3 Weeks ago

- ▶ commit 最多的一周，一屏幕放不下
- ▶ 支持了 musl 的 tls
- ▶ 添加了许多 syscall 的支持，特别是 sys_poll 和 sys_select
- ▶ nc 可以一边读串口一边等待网络
- ▶ 实现了 sys_clone sys_futex 完善了 sys_fork 和 sys_exec
- ▶ 首次支持了 Nginx (3 月 8 日)
- ▶ 首次支持了 Redis (3 月 10 日)

▼ General

Request URL: http://10.0.0.2/

Request Method: GET

Status Code: ● 304 Not Modified

Remote Address: 10.0.0.2:80

Referrer Policy: no-referrer-when-downgrade

▼ Response Headers [view source](#)

Connection: keep-alive

Date: Fri, 08 Mar 2019 10:01:10 GMT

ETag: "0-75"

Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT

Server: nginx/1.14.2

▼ Request Headers [view source](#)

Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,*/*;q=0.8,application/signed-exchange;v=b3

Accept-Encoding: gzip, deflate

Accept-Language: zh-CN,zh;q=0.9,en;q=0.8

```
jiegec@vision2 ~/r/kernel> redis-cli -h 10.0.0.2 set abc def
OK
jiegec@vision2 ~/r/kernel> redis-cli -h 10.0.0.2 get abc
"def"
jiegec@vision2 ~/r/kernel>
```

2 Weeks Ago

- ▶ IXGBE 网卡驱动 82599ES SFP+ “Loopback”
ref: Biscuit
- ▶ 优化万兆网卡性能 3.50Gb/s S 4.44Gb/s R
- ▶ 上真机: VGA 输出, 键盘驱动, 各种修复
- ▶ 映射 PCI BAR 地址空间到用户态, 测试用户态驱动
- ▶ 坑: 爆栈

1 Week Ago

- ▶ 扩展 syscall 支持到更多 ISA
- ▶ 添加 HiFive Unleashed 支持
- ▶ 测试用户态网卡驱动，可以打满 10Gb/s 速度，用 Rust 编写 iperf client
- ▶ 完善文件系统，支持将文件映射的 mmap
- ▶ 可以运行动态链接的程序

Within 1 week

- ▶ 实现了符号链接的支持，可以运行 alpine 的 rootfs 的部分应用
- ▶ CI 测试的修复
- ▶ 用 qcow2 代替 img 格式，可以用 alpine 的包管理器安装应用
- ▶ 重构网络 syscall 实现
- ▶ 可以在 rCore 中运行 GCC 并运行编译出来的程序
- ▶ 实现 AHCI 驱动 ref: Biscuit

```
[ WARN] stat is partial implemented as istat
[ WARN] sys_umask is unimplemented
[ WARN] sys_umask is unimplemented
[ WARN] sys_chmod is unimplemented
    * real-1d in the compiler's search directories.
[Process exited with code 0]
/> printf file specified in the REAL_LD_FILE_NAME configuration macro
[ WARN] sys_clone is calling sys_fork instead, ignoring other args
[ WARN] sys_set_tid_address is unimplemented
[ WARN] sys_ioctl is unimplemented
Built within rCore_oids running 1d using the file name under which
[Process exited with code 0]
/>
```


后续的坑？

前人栽树，后人栽树。 – Harry Chen

前人挖坑，后人把前人推进坑里埋了。 –
Harry Chen

前人用铲子挖坑，后人用混凝土填坑。 –
Dang Fan

太多了

- ▶ 内存管理 kmalloc page table
- ▶ 文件系统 mount procfs devfs sysfs
- ▶ 网络栈 listen backlog
- ▶ 同步互斥 condvar
- ▶ 路由器 IP routing
- ▶ 多 ISA 支持 MIPS