第 5 讲：The Interface of OS

第四节：How to design a Linux kernel interface

陈渝

清华大学计算机系

yuchen@tsinghua.edu.cn

2020 年 3 月 15 日

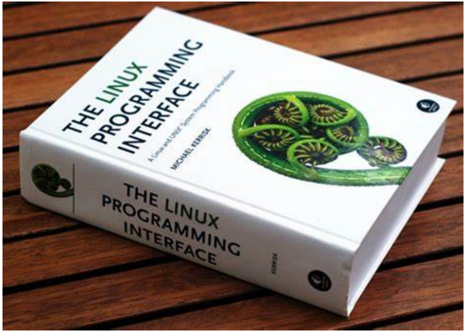
Introduction

FOSDEM 2016

How to design a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels


THE LINUX PROGRAMMING INTERFACE
A Linux and UNIX System Programming Handbook
MICHAEL KERRISK
THE LINUX PROGRAMMING INTERFACE
KERRISK

Introduction

FOSDEM 2016

How to design
a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 1: diverse user usages

try to imagine the ways in which an army of inventive user-space programmers might (ab)use your API

3.5 MQ changes also broke user space in at least two places

- Introduced hard limit of 1024 on queues_max, Fixed by commit f3713fd9c
- Semantics of value exported in /dev/mqueue QSIZE field changed

Introduction

FOSDEM 2016

How to design
a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 2 : unit tests

without unit tests you will screw up someone’ s API

Regressions happen more often than you’ d expect

- Linux 2.6.12 silently changed meaning of fcntl() F_SETOWN
- No longer possible to target signals at specific thread in multithreaded process

Inotify IN_ONESHOT flag

- By design, IN_ONESHOT did not cause an IN_IGNORED event when watch is dropped after one event
- From 2.6.36, IN_ONESHOT does cause

Introduction

FOSDEM 2016

How to design a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 3: Specification, Andrew Morton

Programming is not just an act of telling a computer what to do: it is also an act of telling other programmers what you wished the computer to do. Both are important, and the latter deserves care.

recvmmsg() timeout argument needed a specification; something like:

- timeout is NULL
- timeout points to 0, 0
- timeout points to a structure which is nonzero.
- If, while blocking, the call is interrupted by a signal handler.

Introduction

FOSDEM 2016

How to design a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 4: feedback loop

Strive to shorten worst-case feedback loop. Publicize API design as widely + early as possible.

Ideally, do all of the following before API release:

- Write a detailed specification
- Write example programs that fully demonstrate API
- Email relevant mailing lists and, especially, relevant people, CC linux-api@vger.kernel.org
- write an LWN.net article

Introduction

FOSDEM 2016

How to design a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 5: into real world

Only way to discover design problems in a new nontrivial API is by writing complete, real-world application(s)

Writing a “real” inotify application

- Back story: I thought I understood inotify
- Then I tried to write a “real” application(500 lines of C with (lots of) comments)...
- Written up on LWN (https://lwn.net/Articles/605128/)
- And understood all the work that inotify still leaves you to do

Introduction

FOSDEM 2016

How to design
a Linux kernel interface

Michael Kerrisk
man7.org Training and Consulting
http://man7.org/training/

31 January 2016
Bruxelles / Brussel / Brussels

Moral 6: technical checklist

- New system calls should allow for extensibility.
- Undefined arguments and flags must be zero.
- Syscalls with timeouts should allow absolute timeouts
- Avoid extending multiplexor system calls, etc.