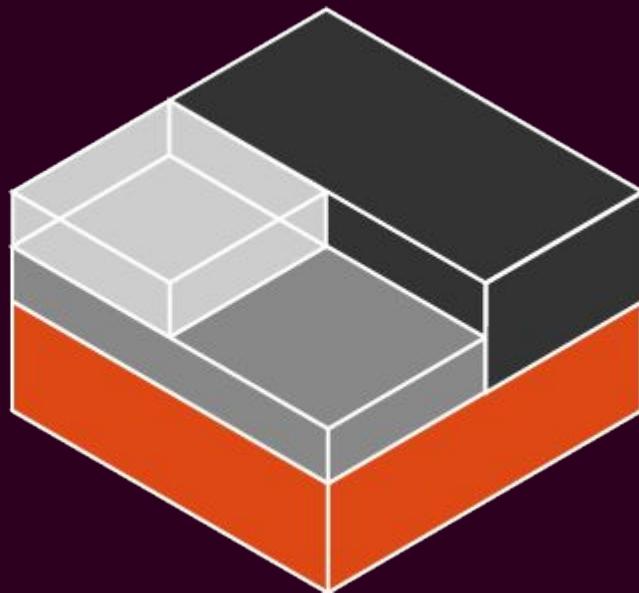


Live migration in LXD

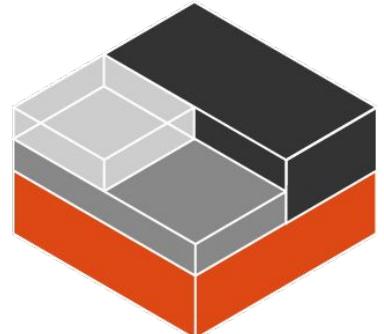
`lxc move foo host2:`



Tycho Andersen, Canonical Ltd.

tycho.andersen@canonical.com
<http://tycho.ws>

Who is this guy?

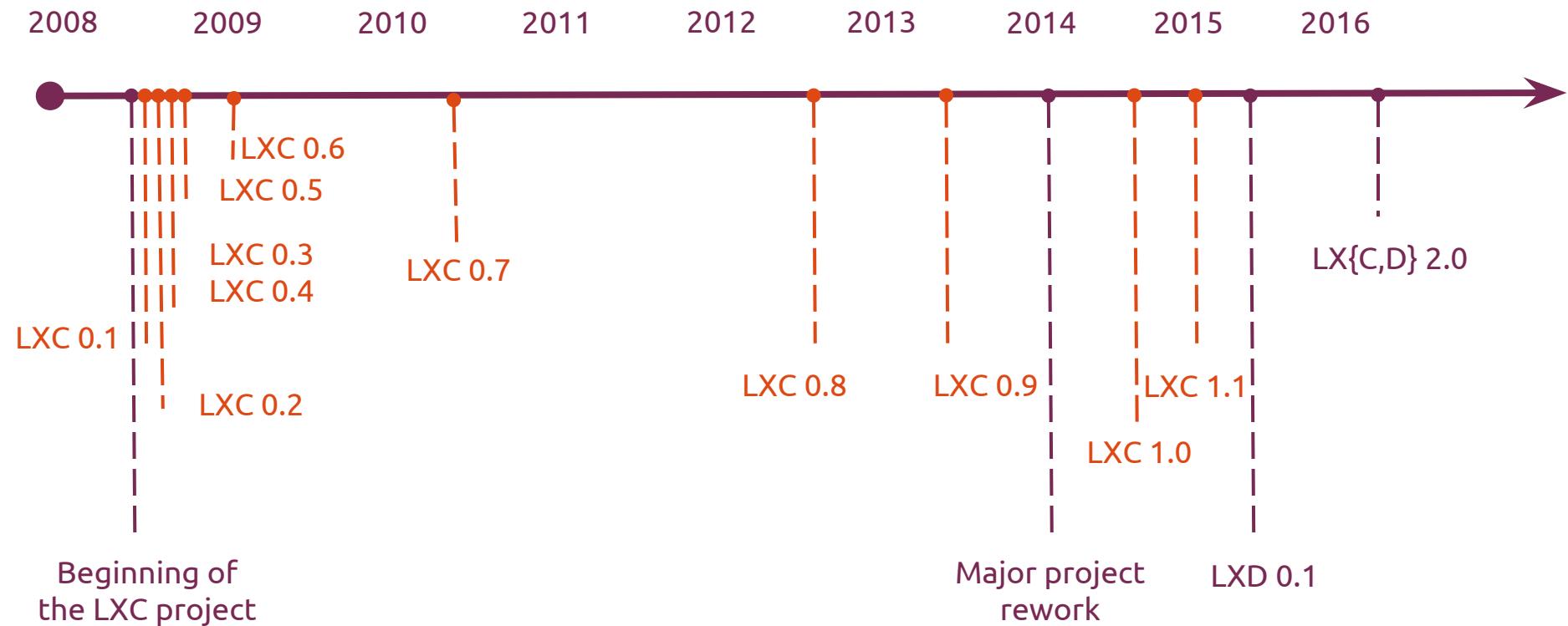
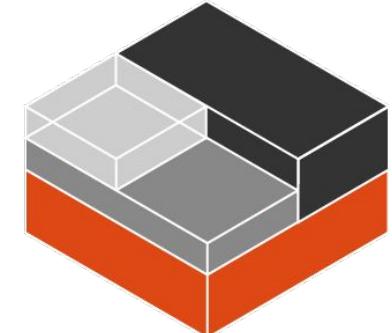


- LXD
- LXC
- CRIU
- Kernel

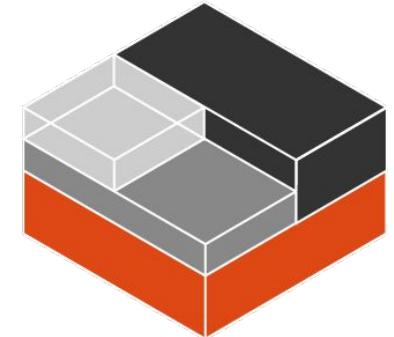


The history of LXC

Over 7 years of Linux system containers



LXD: a modern frontend for LXC

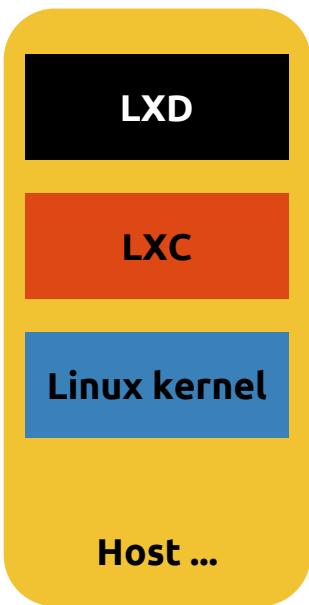
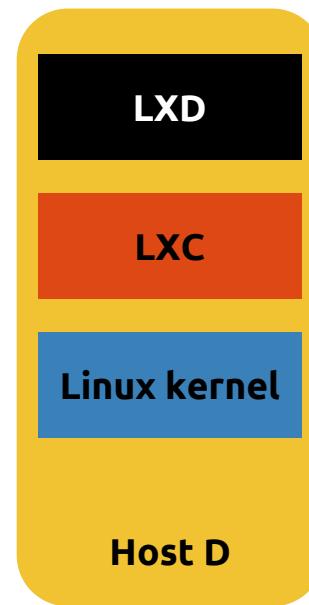
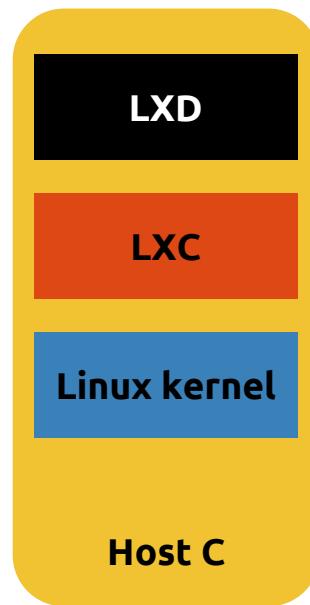
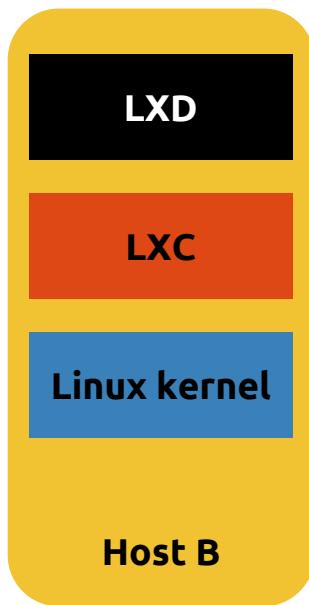
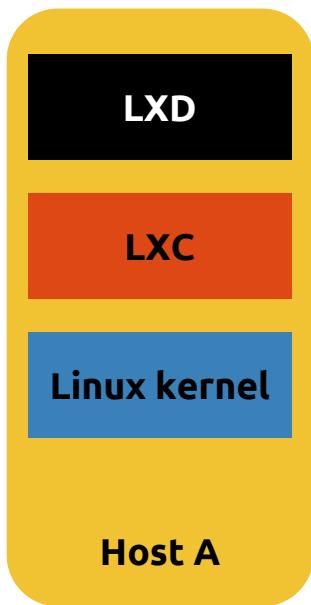


nova-lxd

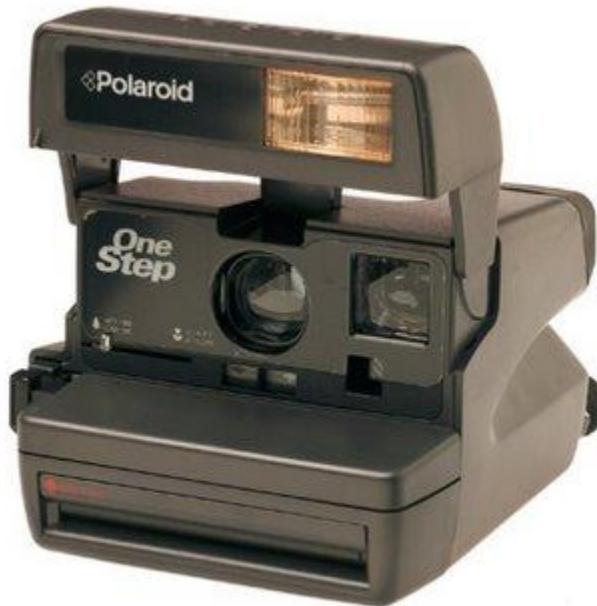
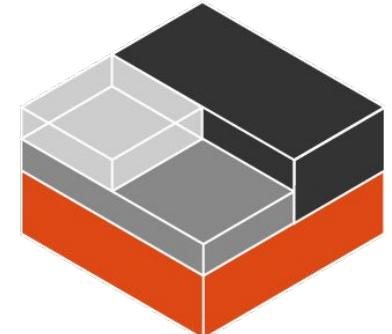
lxc (command line tool)

your own client/script ?

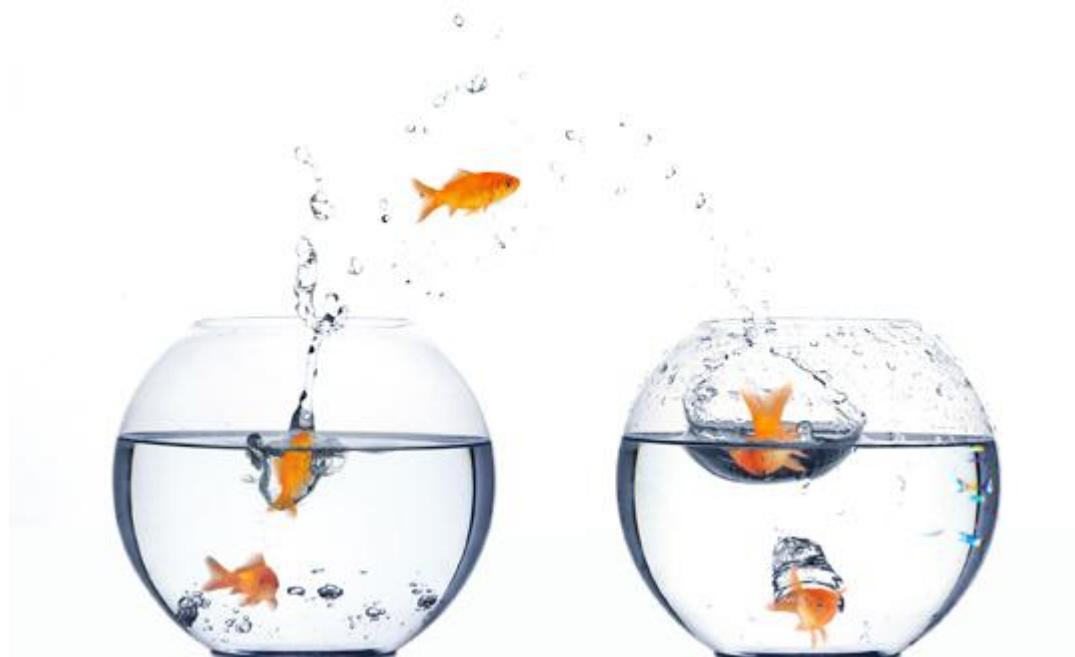
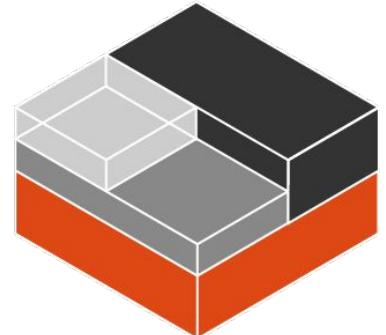
LXD REST API



Hypervisor-y things



`lxc move host1:c1 host2:`

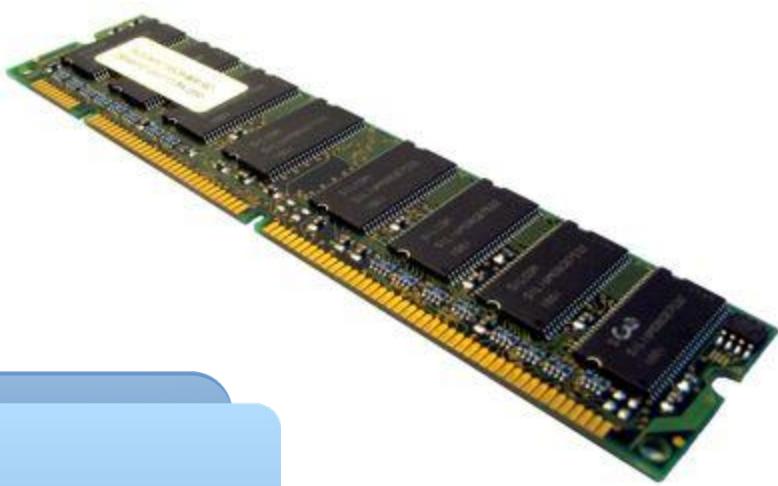
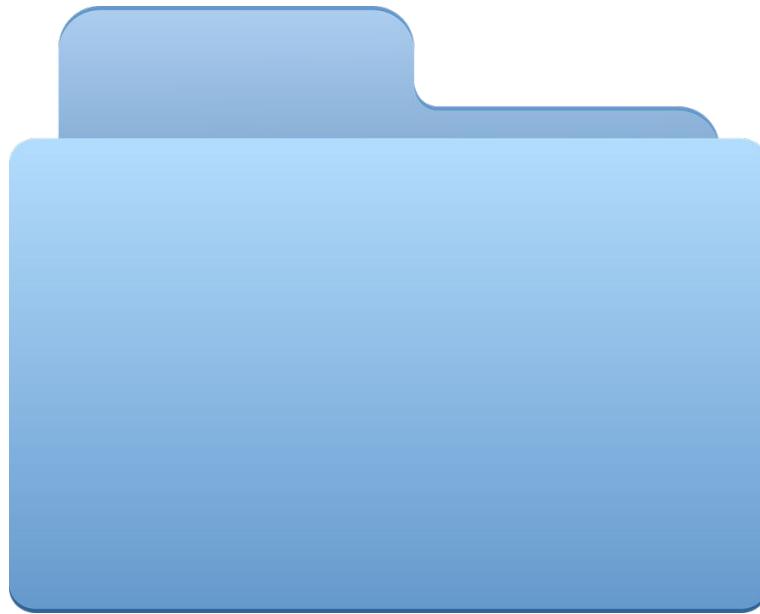


What *actually happens* when you type
`lxc move`?

`lxc move host1:foo host2:`



`lxc move host1:foo host2:`

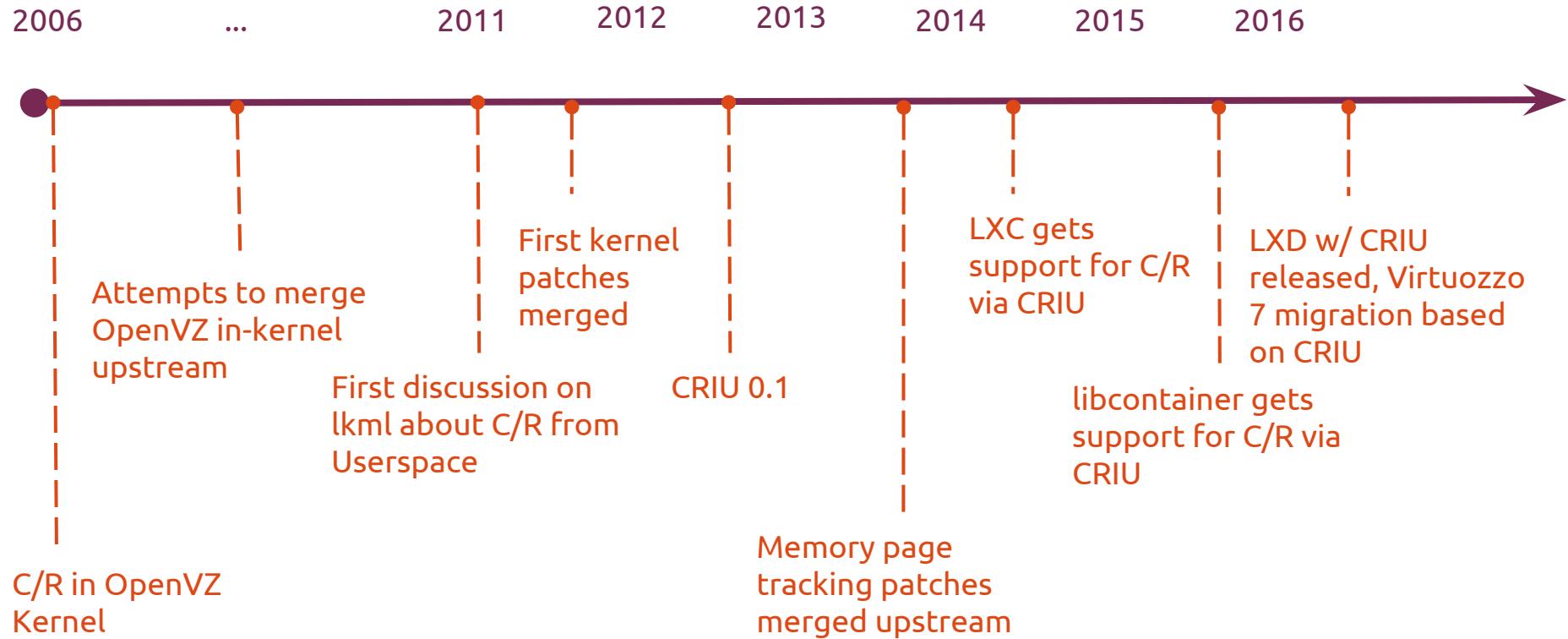
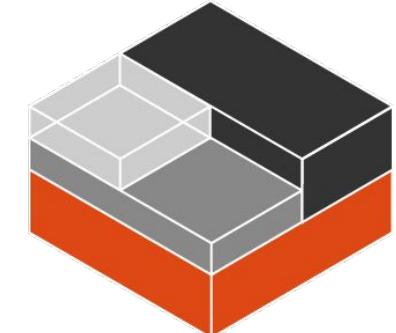




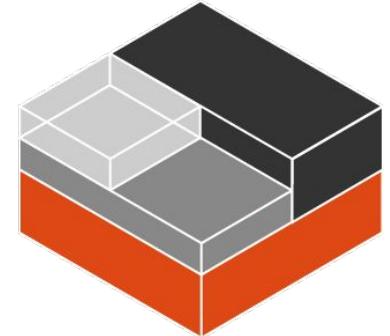


The history of CRIU

Five years of checkpointing!



What's the catch?



“A note on this: this is a project by various mad Russians to perform c/r mainly from userspace, with various oddball helper code added into the kernel where the need is demonstrated... However I'm less confident than the developers that it will all eventually work!”

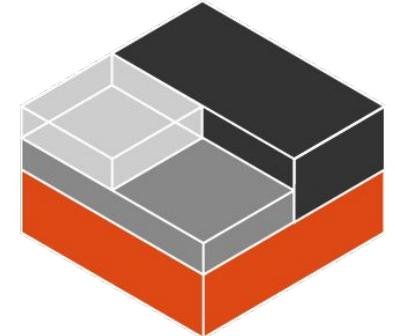
- Linus Torvalds (kernel commit 09946950)

How does it work?

```
[lxc monitor] /var/lib/lxd/containers zesty
\_\_ /sbin/init
    \_\_ /lib/systemd/systemd-journald
    \_\_ /lib/systemd/systemd-udevd
    \_\_ [udevadm] <defunct>
    \_\_ /bin/sh /etc/init.d/apparmor start
        \_\_ systemctl -p LoadState apparmor
    \_\_ /bin/bash /etc/init.d/ebtables start
        \_\_ /bin/bash /etc/init.d/ebtables start
        \_\_ systemctl -p LoadState ebttables
    \_\_ /usr/bin/python3 /usr/bin/cloud-init init
```

How does it work?

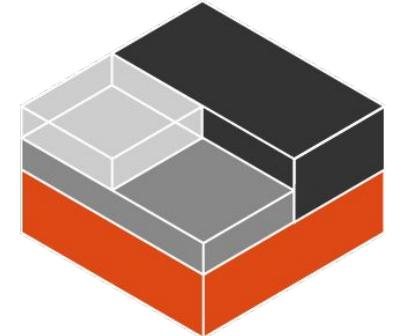
```
[lxc monitor] /var/lib/lxd/containers zesty
\_\_ 6 * fork(); setproctitle("init"); open(); ...
    \_\_ setproctitle("systemd-journald"); bind();...
    \_\_ setproctitle("systemd-udevd"); ...
    \_\_ setproctitle("[udevadm]"); exit();
    \_\_ fork(); setproctitle("sh"); ...
        \_\_ open("/etc/apparmor.d"); mmap(); ...
    \_\_ fork(); setsid(); ...
        \_\_ fork(); setproctitle("bash"); ...
            \_\_ socket(AF_NETLINK, ...); setuid(); ...
    \_\_ setproctitle("cloud-init"); ...
```



“This is not an enterprise feature. It's a promise one cannot keep. We will not add code to systemd that works often but not always, and CRIU is certainly of that kind.”

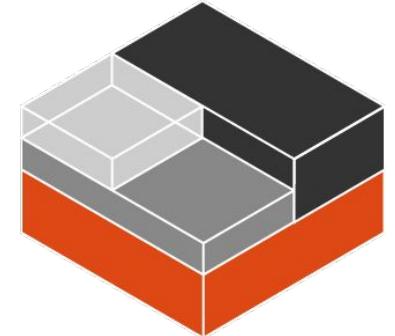
- Lennart Potterink (*systemd-devel*, 2015)





“This gives me the creeps.”

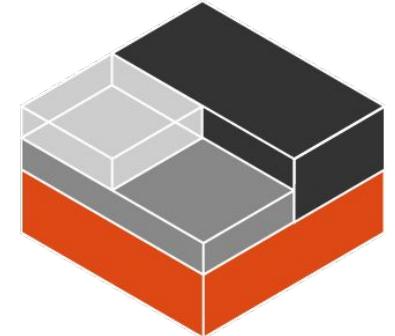
- Kees Cook (seccomp maintainer, lkml, 2015)



“Can we refuse dumping selinux labels...until we understand how to properly do it?”
- Pavel Emelyanov (CRIU list, May 2015)

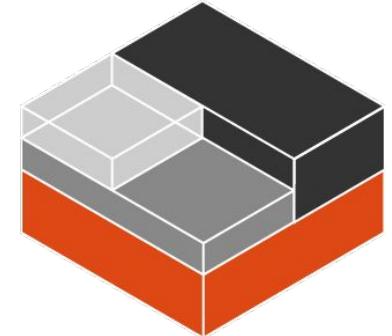


Smoke and mirrors!



Coming down the pike

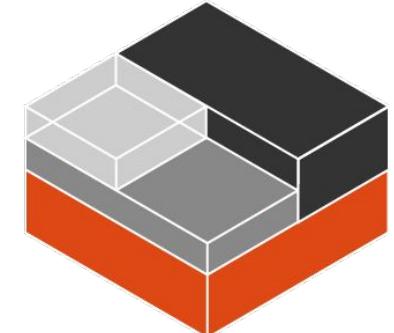
- task_diag work by Andrei Vagan
- AF_UNIX and SCM_*
- userfaultfd (post-copy) by Adrain Reber && Mike Rapoport
- More LSMs and PTRACE_O_SUSPEND_LSM
- netlink
- P.haul for go



Correct and Fast

Pick two

Administrivia



→ LXD

- ◆ Current release 2.5
- ◆ 2.6 targeted for November 2016
- ◆ Monthly release cadence
- ◆ <https://linuxcontainers.org>
- ◆ <https://github.com/lxc/lxd>

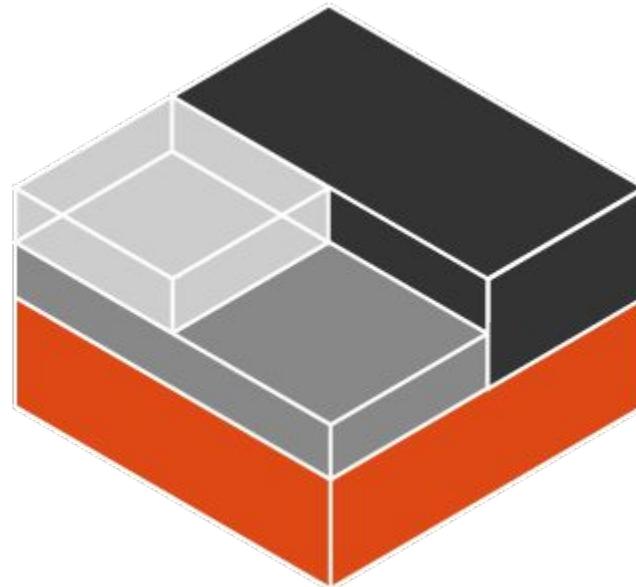
→ CRIU

- ◆ Current stable release 2.7 “Rubber Owl”
- ◆ 2.8 Targeted for November 2016
- ◆ Monthly release cadence
- ◆ <http://criu.org>
- ◆ <https://github.com/xemul/criu>

Tycho Andersen, Canonical Ltd.

tycho.andersen@canonical.com
<http://tycho.ws>

<https://linuxcontainers.org/lxd>
<https://github.com/lxc/lxd>



Questions?

CANONICAL