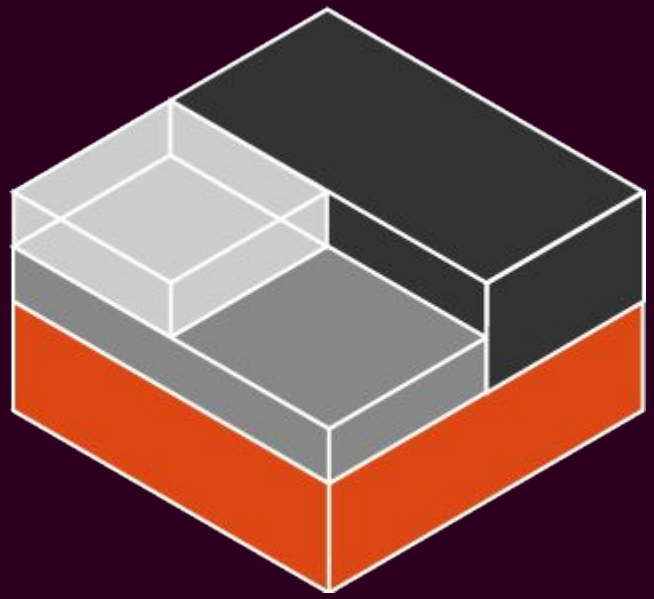
ubuntu

live migration of linux containers

lxc move foo host2:

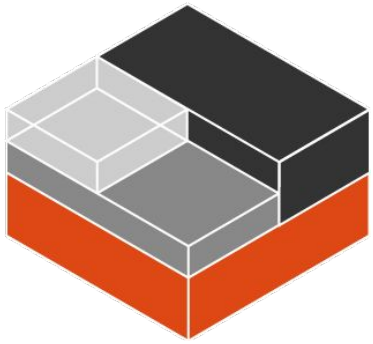


Tycho Andersen, Canonical Ltd.

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

CANONICAL

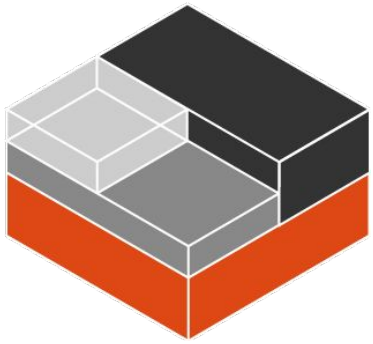
Who is this guy?

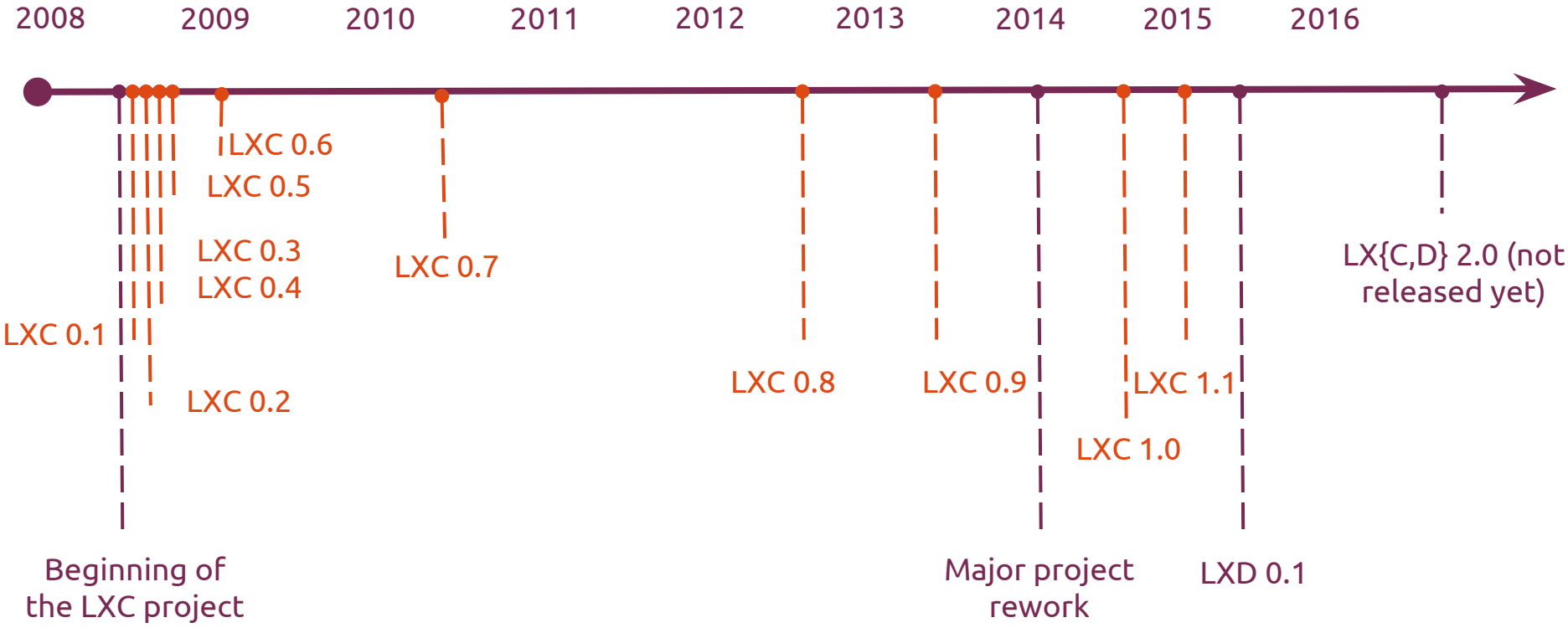


- ➔ LXD
- ➔ LXC
- ➔ CRIU
- ➔ Kernel

The history of LXC

Over 7 years of Linux system containers



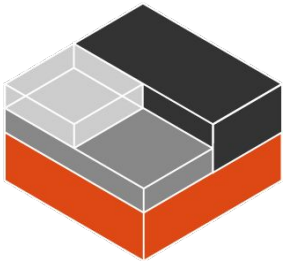



LEGO

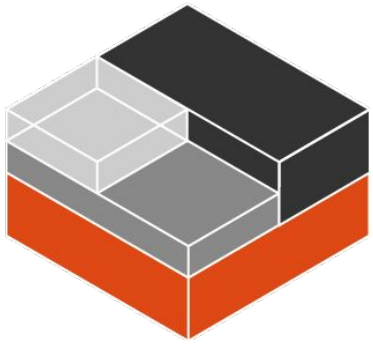

AUSTRALIA
MELBOURNE
UBS
EXTRA FORMATION LAP
MELBOURNE
ROLEX


FICHET
PARIS

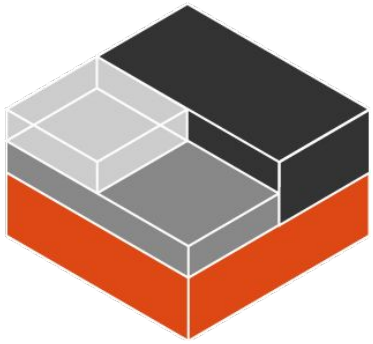
Microsoft
Hyper-V



Hypvervisor-y things



LXD: the container lighter-visor

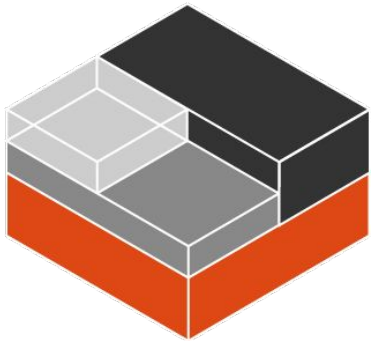


nova-compute-lxd | lxc (command line tool) | your own client/script ?

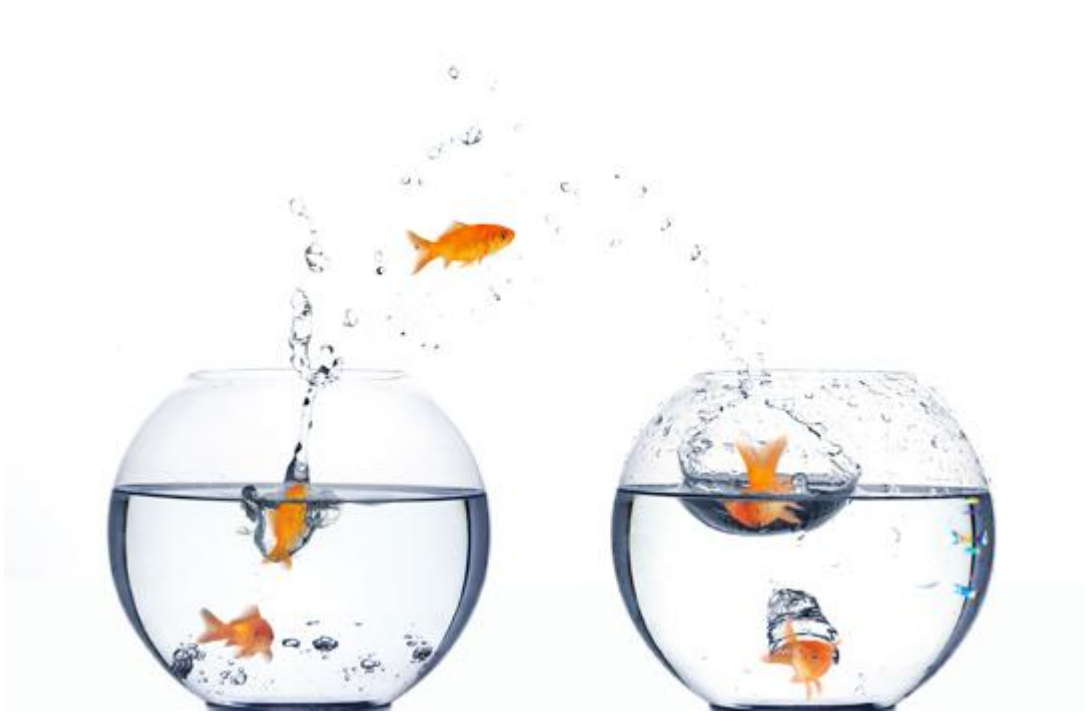
LXD REST API

Host A	Host B	Host C	Host D	Host ...
LXD	LXD	LXD	LXD	LXD
LXC	LXC	LXC	LXC	LXC
Linux kernel	Linux kernel	Linux kernel	Linux kernel	Linux kernel

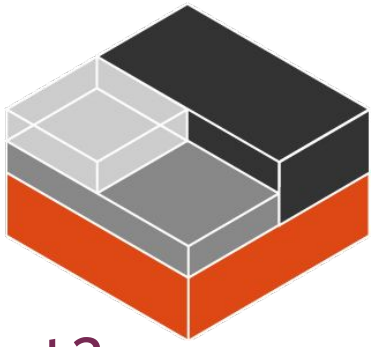
Hypvervisor-y things



lxc move host1:c1 host2:



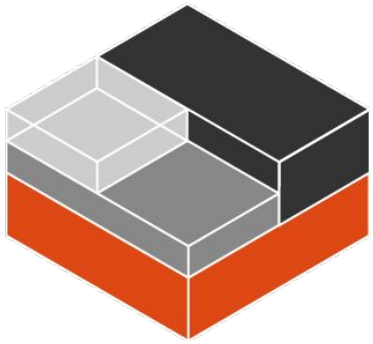
lxc move host1:c1 host2:

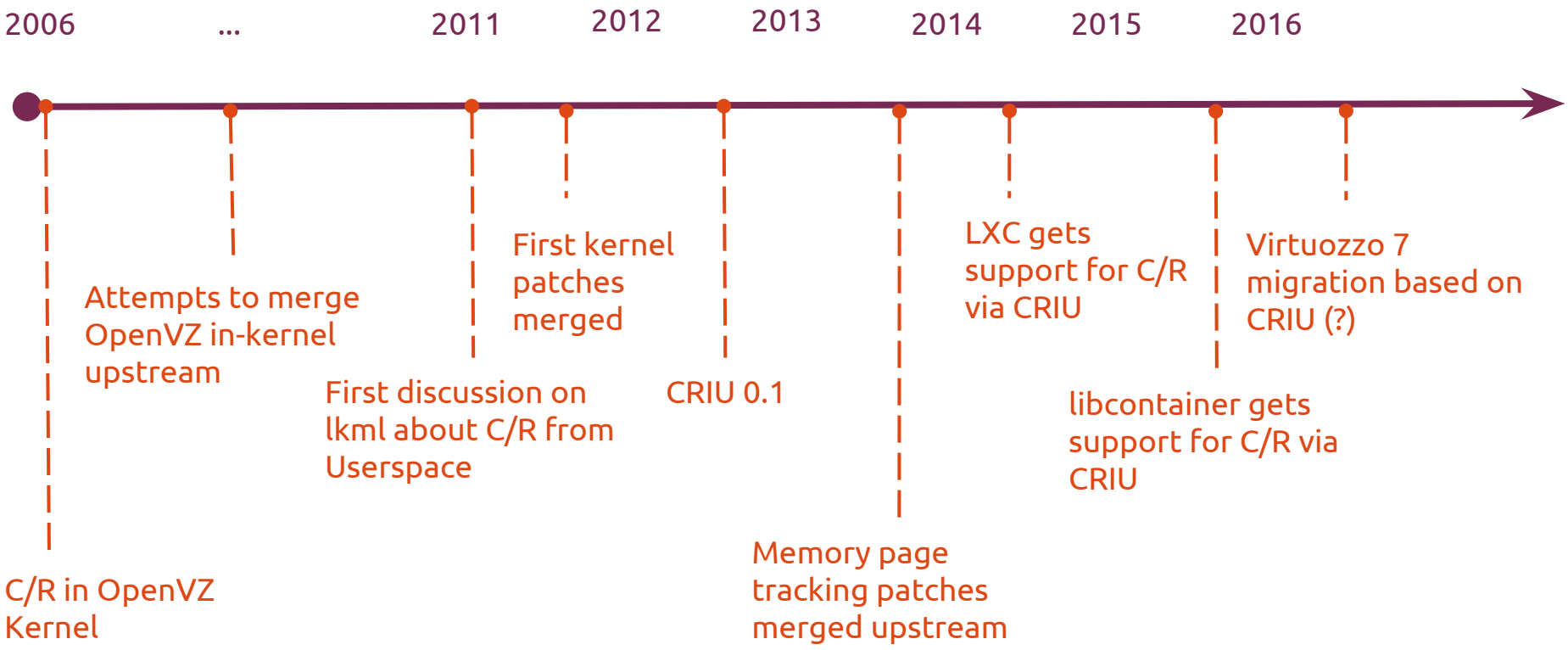


- host1 negotiates three “channels” with host2
  - control
  - filesystem
  - container process state
- Using a tool called CRIU for process state

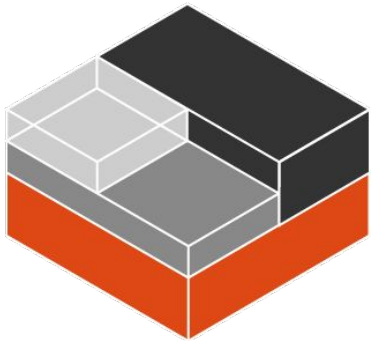
The history of CRIU

Five years of checkpointing!



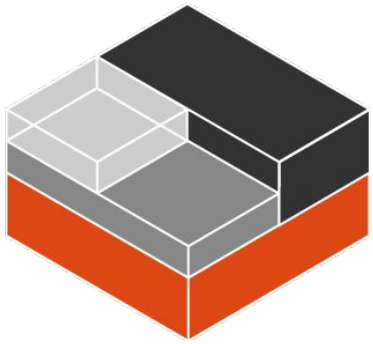


What’s the catch?



“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)

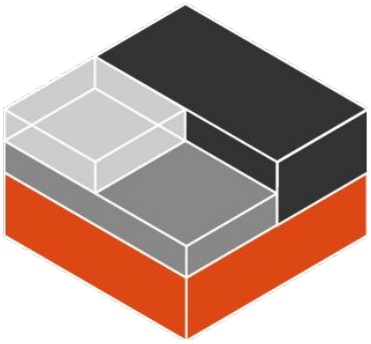


“This is not an enterprise feature. It's a promise one cannot keep. We will not add code to systemd that works often but not always, and CRIU is certainly of that kind.”

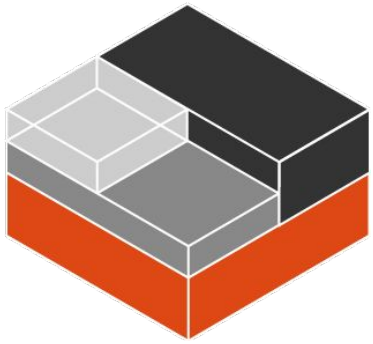
- Lennart Pottering (systemd-devel, 2015)


HEINZ
EST 1869
TOMATO
KETCHUP
57 VARIETIES
GROWN NOT MADE
NET WT 20 OZ (1 LB 4 OZ) 567g

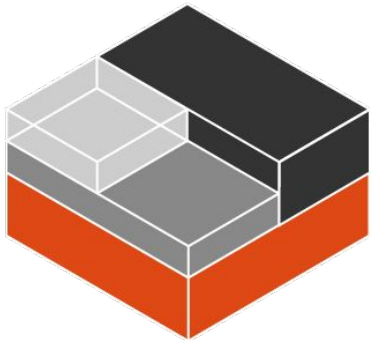
Smoke and mirrors!



Security

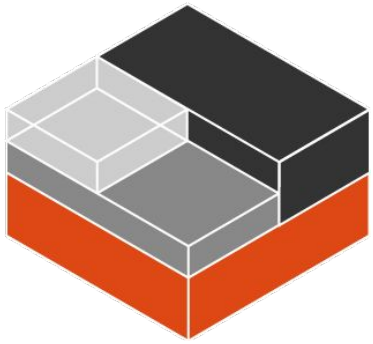


- ➔ cgroups
- ➔ apparmor, ~~selinux~~, etc.
- ➔ seccomp (STRICT, FILTER)
- ➔ user namespaces



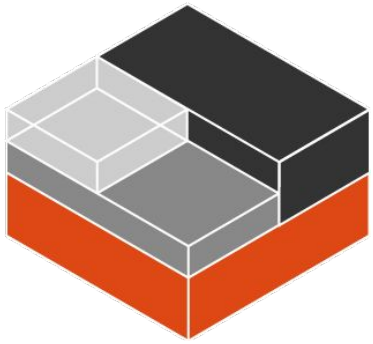
“Can we refuse dumping selinux labels...until we understand how to properly do it?”

- Pavel Emelyanov (CRIU list, May 2015)



“This feature gives me the creeps.”

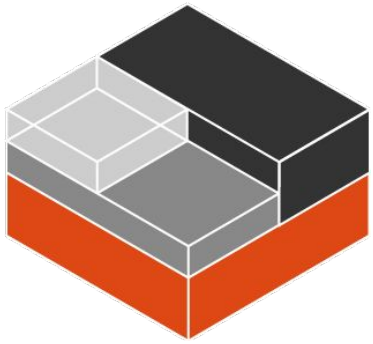
- Kees Cook (seccomp maintainer, lkml, 2015)



Correct and Fast

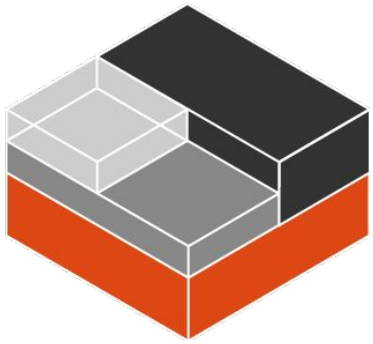
Pick two

Making Migration Fast



- Three channels
  - control
  - filesystem specific
  - memory state specific
- Filesystems:
  - btrfs, LVM, ZFS, (swift, nfs?), etc.
  - rsync between incompatible hosts
- Memory state:
  - Stop the world
  - Iterative incremental transfer (via p.haul)

Administrivia

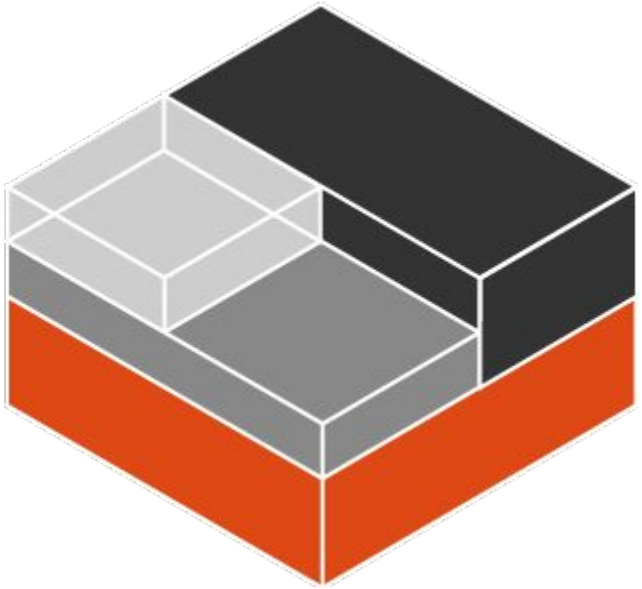


- LXD
  - Current release 2.0.0beta1
  - 2.0 targeted for February 2016
  - Two week release cadence
  - LXC+LXD 2.0s will land in Xenial (as will cgns)
  - https://linuxcontainers.org
  - https://github.com/lxc/lxd
- CRIU
  - Current stable release 1.8
  - Three month release cadence
  - http://criu.org
  - 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?