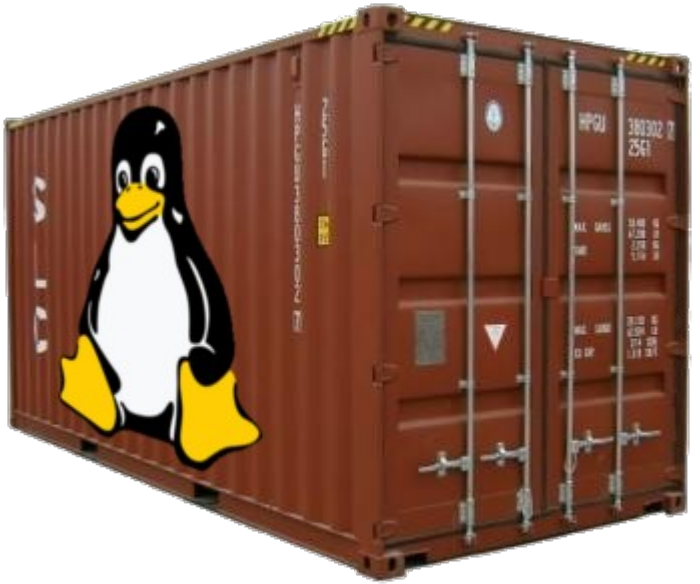
Virtualization which isn't
LXC (Linux Containers)



Dobrica Pavlinušić
http://blog.rot13.org
DORS/CLUC, Zagreb, 2011-05-16

Content

- Virtualizations
  - Vserver, Xen, OpenVZ, sVirt, **LXC**
  - KVM, VirtualBox, VMWare
- cgroup
- Linux Containers

Virtualization overview

- Xen
  - Separate host i guest kernel (dom0, domU)
  - Not upstream, massive duplication of kernel code
- Linux Vserver, OpenVZ (Virtuozzo), sVirt (SELinux based)
  - Single kernel, out-of-tree patches
- **Linux Containers - LXC**
  - chroot on steroids, based on cgroup Linux support
  - Part of standard kernel, based on things you already know!
- Full-system virtualization: KVM, VirtualBox, VMWare
  - But you can run LXC **inside them!** (e.g. EC2)

cgroup

- Process namespace in kernel
  - Devices (even X11 in LXC!)
  - CPU (sched, cpu account, cpuset) - NUMA
  - Memory (not in Debian's kernel)
  - Block I/O scheduling, limits
- Linus' 2.6.38 magic patch
  - Setsid create new scheduler entry
- Used by Google Chrome, systemd...

Linux containers - LXC

```
dpavlin@klin:~$ lxc-checkconfig
Kernel config /proc/config.gz not found, looking in other places...
Found kernel config file /boot/config-2.6.38-2-686
--- Namespaces ---
Namespaces: enabled
Utsname namespace: enabled
Ipc namespace: enabled
Pid namespace: enabled
User namespace: enabled
Network namespace: enabled
Multiple /dev/pts instances: enabled

--- Control groups ---
Cgroup: enabled
Cgroup namespace: enabled
Cgroup device: enabled
Cgroup sched: enabled
Cgroup cpu account: enabled
Cgroup memory controller: missing
Cgroup cpuset: enabled

--- Misc ---
Veth pair device: enabled
Macvlan: enabled
Vlan: enabled
File capabilities: missing
```

```
dpavlin@klin:/usr/bin$ ls lxc-*
lxc-checkconfig
lxc-execute
lxc-start
lxc-stop
lxc-info
lxc-console
lxc-create
lxc-destroy
lxc-ls
lxc-ps
lxc-netstat
lxc-restart
lxc-cgroup
lxc-freeze
lxc-kill
lxc-monitor
lxc-setcap
lxc-setuid
lxc-unfreeze
lxc-unshare
lxc-version
Lxc-wait
lxc-attach
lxc-checkpoint
```

LXC: Network

- veth
  - Bridge on host, (virtual) device inside container
- vlan
  - Select packets by IP address
- macvlan
  - Select packets by MAC address
- phys
  - Move host interface inside container (routing fun!)
- Empty
  - Only loopback

LXC: limit resources

- Cores
  - `lxc.cgroup.cpuset.cpus=1,2,3`
- CPU share
  - `lxc.cgroup.cpu.shares=1024 # default`
- Memory usage (!Debian)
  - `lxc.cgroup.memory.limit_in_bytes = 256M`
  - `lxc.cgroup.memory.memsw.limit_in_bytes = 1G`
- Disk (blkio)
  - Disk space – standard LVM, quota...
  - `echo 100 > /cgroup/disk1/blkio.weight # XXX < 1000 !`
  - `echo "3:0 1048576" > /cgroup/disk1/blkio.throttle.read_bps_device`

Start LXC container

- Start single process in container
  - `lxc-execute -n container -- /bin/bash`
- Whole operating system
  - Mounting filesystems, etc from config file
  - Application is /bin/init
  - `lxc-start -n container`
  - `lxc-console -n container`
  - `lxc-stop -n container`

Templates: lxc-create

```
# /usr/lib/lxc/templates/

export MIRROR=http://192.168.1.20:3142/ftp.debian.org
export SUITE=lenny

cat <<_EOF_ > /tmp/lenny.conf
lxc.network.type=veth
lxc.network.link=br0
lxc.network.flags=up
EOF

t61p:~# lxc-create -n lenny -t debian -f /tmp/lenny.conf
```

Container overview

- /var/lib/lxc/container/config
- Familiar commands
  - lxc-ls
  - lxc-info
  - lxc-ps
  - lxc-netstat
- htop --enable-group > r192
- /proc inside contauner isn't fully isolated!
  - Depends on namespace support in kernel

Under construction

- Still not in: Linux 2.6.38.2
- lxc-attach
  - Attach process (bash) inside running container
  - Needed to set default route outside container
- lxc-checkpoint
  - Similar to lxc-(un)freeze with checkpoint to disk
  - https://ckpt.wiki.kernel.org/

LXC summary

- Isolate
  - one application – lxc-execute
  - whole OS – lxc-start
- use templates (lxc-create)
- Familiar Linux networking (bridges)
- Limiting features varies (kernel config/version)
- Ready to use today!