

Using Containers Without Risking Your *aas (Canonical)

Serge Hallyn, Scott Moser

Canonical, Inc

serge.hallyn@ubuntu.com, scott.moser@canonical.com

November 2, 2014

Overview

- 1 Quick Introduction to Containers
- 2 State of Containers Prior to User namespaces
- 3 User namespaces
- 4 Graphical Demo

Linux Containers?

- operating system-level virtualization method for running multiple isolated Linux systems (containers) on a single control host.
- "chroot on steroids"
- "it's like bsd jails" (or solaris zones)
- from the inside looks like a vm
- from the outside looks like processes

Containers prior to user namespaces

Namespaces

- *id* → *resource* mapping
 - Prevent resource access by not providing a handle
 - i.e. pid 1 is not global init
 - /etc/shadow not accessible
- Many leaks (/proc/pid/fd/N)

Control groups

- 1 Resource limits and accounting
- 2 Limit device access
- 3 If root, re-mount cgroups and change/escape limits.

Capabilities bounding set

- 1 Limit privs of root in container
- 2 Root still owns most host files
- 3 http://www.sevagas.com/IMG/pdf/exploiting_capabilities_the_dark_side.pdf
- 4 Prevents useful things like tmpfs mounts

LSMs

- 1 Paper over the (huge) remaining holes
- 2 i.e. prevent `/proc/sys/*` writing, etc
- 3 "Safe from accidental damage by container root"
- 4 People always want unsafe exceptions
- 5 Lack of policy nesting limits use *in* containers

Seccomp

- 1 Prevent use of some syscalls
- 2 Reduce exposed kernel surface
- 3 Hard to do generally

① Nevertheless

- ① Root in container is still root on host
- ② Any leak = game over
- ③ Answer: "Wait for user namespaces"

Demo Time [sort of].

- 1 Ubuntu 14.10 instance with hostname 'lxc-host'.
- 2 2 users (elsa, anna) are each configured to run lxc unprivileged.
- 3 'showinfo': simple shell filter to 'find' or 'ps' or 'grep'.
- 4 'mywait': Very Exciting. Run it, it prints its pid, uid, gid. Then creates a file named 'sleeper-user@hostname' and sleeps forever. copied into each container's /usr/local/bin.

Host Processes / Users.

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid                gid                mode
sleeper-elsa@lxc-host.info  uid=elsa          gid=elsa          mode=664
sleeper-root@lxc-host.info  uid=root          gid=root          mode=644
sleeper-anna@lxc-host.info  uid=anna          gid=anna          mode=664
sleeper-root@anna-c1.info   uid=2000000       gid=2000000      mode=664
sleeper-root@anna-c2.info   uid=2100000       gid=2100000      mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000       gid=3001000      mode=644
sleeper-root@elsa-c2.info   uid=3100000       gid=3100000      mode=664
root@lxc-host# showinfo ps
command            uid                gid                pid                users
/sbin/init         root               0                  1                 4026531837
/sbin/init         2000000           2000000           6900              4026532171
/sbin/init         2100000           2100000           7704              4026532294
/sbin/init         3000000           3000000           8590              4026532233
/sbin/init         3100000           3100000           9340              4026532356
sleeper-anna@lxc-host  anna              1001              11826             4026531837
sleeper-root@anna-c1  2000000           2000000           12177             4026532171
sleeper-root@anna-c2  2100000           2100000           12426             4026532294
sleeper-elsa@lxc-host  elsa              1002              13113             4026531837
sleeper-ubuntu@elsa-c1 3001000           3001000           14280             4026532233
sleeper-root@elsa-c2  3100000           3100000           14535             4026532356
sleeper-root@lxc-host  root              0                  15506             4026531837
root@lxc-host#

elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002

anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001

elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000

anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0

root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0

root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0

root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535
root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
root@lxc-host#
```


LXC Containers and Configuration

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid                gid                mode
sleeper-elsa@lxc-host.info  uid=elsa          gid=elsa          mode=664
sleeper-root@lxc-host.info  uid=root          gid=root          mode=644
sleeper-anna@lxc-host.info  uid=anna          gid=anna          mode=664
sleeper-root@anna-c1.info   uid=2000000       gid=2000000       mode=664
sleeper-root@anna-c2.info   uid=2100000       gid=2100000       mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000       gid=3001000       mode=644
sleeper-root@elsa-c2.info   uid=3100000       gid=3100000       mode=664
root@lxc-host# showinfo ps
command  uid  gid  pid  users
/sbin/init  root  0  1  4026531837
/sbin/init  2000000  2000000  6900  4026532171
/sbin/init  2100000  2100000  7704  4026532294
/sbin/init  3000000  3000000  8590  4026532233
/sbin/init  3100000  3100000  9340  4026532356
sleeper-anna@lxc-host  anna  1001  11826  4026531837
sleeper-root@anna-c1  2000000  2000000  12177  4026532171
sleeper-root@anna-c2  2100000  2100000  12426  4026532294
sleeper-elsa@lxc-host  elsa  1002  13113  4026531837
sleeper-ubuntu@elsa-c1  3001000  3001000  14280  4026532233
sleeper-root@elsa-c2  3100000  3100000  14535  4026532356
sleeper-root@lxc-host  root  0  15506  4026531837
root@lxc-host#

root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535

root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000

root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000

root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
root@lxc-host#

elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002

anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001

elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000

anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0

root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0

root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0

u* 14.10 0:-* 20m 0.00 2.0GHz 2.0G15% 2014-10-30 18:25:41
```

Anna's containers: anna-c1, anna-c2

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid                gid                mode
sleeper-elsa@lxc-host.info  uid=elsa          gid=elsa          mode=664
sleeper-root@lxc-host.info  uid=root          gid=root          mode=644
sleeper-anna@lxc-host.info  uid=anna          gid=anna          mode=664
sleeper-root@anna-c1.info   uid=2000000       gid=2000000      mode=664
sleeper-root@anna-c2.info   uid=2100000       gid=2100000      mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000       gid=3001000      mode=644
sleeper-root@elsa-c2.info   uid=3100000       gid=3100000      mode=664
root@lxc-host# showinfo ps
command             uid                gid                pid                usersn
/sbin/init          root               0                  1                  4026531837
/sbin/init          2000000           2000000           6900               4026532171
/sbin/init          2100000           2100000           7704               4026532294
/sbin/init          3000000           3000000           8590               4026532233
/sbin/init          3100000           3100000           9340               4026532356
sleeper-anna@lxc-host  anna              1001              11826              4026531837
sleeper-root@anna-c1  2000000           2000000           12177              4026532171
sleeper-root@anna-c2  2100000           2100000           12426              4026532294
sleeper-elsa@lxc-host  elsa              1002              13113              4026531837
sleeper-ubuntu@elsa-c1 3001000           3001000           14280              4026532233
sleeper-root@elsa-c2  3100000           3100000           14535              4026532356
sleeper-root@lxc-host  root              0                  15506              4026531837
root@lxc-host#
root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535
root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
root@lxc-host#
elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002
anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001
elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000
anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0
root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0
root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0
u® 14.10 0:-* 20m 0.00 2.0GHz 2.0G15% 2014-10-30 18:25:41
```

Anna's containers: anna-c1, anna-c2

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid                gid                mode
sleeper-elsa@lxc-host.info  uid=elsa          gid=elsa          mode=664
sleeper-root@lxc-host.info  uid=root          gid=root          mode=644
sleeper-anna@lxc-host.info  uid=anna          gid=anna          mode=664
sleeper-root@anna-c1.info   uid=2000000       gid=2000000      mode=664
sleeper-root@anna-c2.info   uid=2100000       gid=2100000      mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000       gid=3001000      mode=644
sleeper-root@elsa-c2.info   uid=3100000       gid=3100000      mode=664
root@lxc-host# showinfo ps
command             uid                gid                pid         users
/sbin/init          root               0                  1           4026531837
/sbin/init          2000000           2000000           6900        4026532171
/sbin/init          2100000           2100000           7704        4026532294
/sbin/init          3000000           3000000           8590        4026532233
/sbin/init          3100000           3100000           9340        4026532356
sleeper-anna@lxc-host  anna              1001              11826       4026531837
sleeper-root@anna-c1  2000000           2000000           12177       4026532171
sleeper-root@anna-c2  2100000           2100000           12426       4026532294
sleeper-elsa@lxc-host  elsa              1002              13113       4026531837
sleeper-ubuntu@elsa-c1 3001000           3001000           14280       4026532233
sleeper-root@elsa-c2  3100000           3100000           14535       4026532356
sleeper-root@lxc-host  root              0                  15506       4026531837
root@lxc-host#

elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002

elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000

root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0

anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001

anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0

root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0

root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535
root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
```

Elsa's Containers: elsa-c1, elsa-c2

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid                gid                mode
sleeper-elsa@lxc-host.info  uid=elsa          gid=elsa          mode=664
sleeper-root@lxc-host.info  uid=root          gid=root          mode=644
sleeper-anna@lxc-host.info  uid=anna          gid=anna          mode=664
sleeper-root@anna-c1.info   uid=2000000       gid=2000000       mode=664
sleeper-root@anna-c2.info   uid=2100000       gid=2100000       mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000       gid=3001000       mode=644
sleeper-root@elsa-c2.info   uid=3100000       gid=3100000       mode=664
root@lxc-host# showinfo ps
command             uid                gid                pid        usersn
/sbin/init          root               0                  1          4026531837
/sbin/init          2000000           2000000           6900       4026532171
/sbin/init          2100000           2100000           7704       4026532294
/sbin/init          3000000           3000000           8590       4026532233
/sbin/init          3100000           3100000           9340       4026532356
sleeper-anna@lxc-host  anna              1001              11826      4026531837
sleeper-root@anna-c1  2000000           2000000           12177      4026532171
sleeper-root@anna-c2  2100000           2100000           12426      4026532294
sleeper-elsa@lxc-host  elsa              1002              13113      4026531837
sleeper-ubuntu@elsa-c1 3001000           3001000           14280      4026532233
sleeper-root@elsa-c2  3100000           3100000           14535      4026532356
sleeper-root@lxc-host  root              0                  15506      4026531837
root@lxc-host#

root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535
root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
root@lxc-host#

elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002

anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001

elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000

anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0

root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0

root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0

u® 14.10 0:-* 20m 0.00 2.0GHZ 2.0G15% 2014-10-30 18:25:41
```

Elsa's Containers: elsa-c1, elsa-c2

```
ubuntu@lxc-host (10.5.0.232) - byobu
root@lxc-host# showinfo files
fname                uid          gid          mode
sleeper-elsa@lxc-host.info  uid=elsa    gid=elsa    mode=664
sleeper-root@lxc-host.info  uid=root    gid=root    mode=644
sleeper-anna@lxc-host.info  uid=anna    gid=anna    mode=664
sleeper-root@anna-c1.info   uid=2000000 gid=2000000 mode=664
sleeper-root@anna-c2.info   uid=2100000 gid=2100000 mode=664
sleeper-ubuntu@elsa-c1.info uid=3001000 gid=3001000 mode=644
sleeper-root@elsa-c2.info   uid=3100000 gid=3100000 mode=664
root@lxc-host# showinfo ps
command  uid  gid  pid  users
/sbin/init  root  0  1  4026531837
/sbin/init  2000000 2000000 6900 4026532171
/sbin/init  2100000 2100000 7704 4026532294
/sbin/init  3000000 3000000 8590 4026532233
/sbin/init  3100000 3100000 9340 4026532356
sleeper-anna@lxc-host  anna  1001  11826 4026531837
sleeper-root@anna-c1  2000000 2000000 12177 4026532171
sleeper-root@anna-c2  2100000 2100000 12426 4026532294
sleeper-elsa@lxc-host  elsa  1002  13113 4026531837
sleeper-ubuntu@elsa-c1 3001000 3001000 14280 4026532233
sleeper-root@elsa-c2  3100000 3100000 14535 4026532356
sleeper-root@lxc-host  root  0  15506 4026531837
root@lxc-host#

root@lxc-host# showinfo config
anna-c1/config:lxc.id_map = u 0 2000000 65535
anna-c1/config:lxc.id_map = g 0 2000000 65535
anna-c2/config:lxc.id_map = u 0 2100000 65535
anna-c2/config:lxc.id_map = g 0 2100000 65535
elsa-c1/config:lxc.id_map = u 0 3000000 65535
elsa-c1/config:lxc.id_map = g 0 3000000 65535
elsa-c2/config:lxc.id_map = u 0 3100000 65535
elsa-c2/config:lxc.id_map = g 0 3100000 65535
root@lxc-host# cat /etc/subuid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# cat /etc/subgid
anna:2000000:1000000
elsa:3000000:1000000
root@lxc-host# moreinfo
https://github.com/hallyn/ods-2014-usersn
cubswin:)
root@lxc-host#

elsa@lxc-host$ mywait
[sleeper-elsa@lxc-host] pid=13109 uid=1002 gid=1002

anna@lxc-host$ mywait
[sleeper-anna@lxc-host] pid=11822 uid=1001 gid=1001

elsa@lxc-host$ lxc_attach elsa-c1
root@elsa-c1# sudo -Hu ubuntu mywait
[sleeper-ubuntu@elsa-c1] pid=597 uid=1000 gid=1000

anna@lxc-host$ lxc_attach anna-c1
root@anna-c1# mywait
[sleeper-root@anna-c1] pid=592 uid=0 gid=0

root@elsa-c2# mywait
[sleeper-root@elsa-c2] pid=597 uid=0 gid=0

root@anna-c2# mywait
[sleeper-root@anna-c2] pid=593 uid=0 gid=0

u* 14.10 0:-* 20m 0.00 2.0GHZ 2.0G15% 2014-10-30 18:25:41
```

Goals

- 1 Uid separation
 - 1 $c1.500 \neq c2.500$
 - 2 Separate access controls (kill, open, etc)
 - 3 Separate accounting, limits
- 2 Container root privileged over container
 - 1 uids
 - 2 network
 - 3 etc
- 3 Container root has no privilege outside of container
 - 1 Root in container as safe as unpriv user on host
 - 2 Safe for use by untrusted users
- 4 Able to be nested

User namespace design

- 1 By Eric Biederman
- 2 Uids map 1-1 to kuid
 - 1 Translated at kernel-user boundary
 - 2 Default mapping 0-4294967295:0-4294967295
 - 3 Unmapped userids show up as -1, has 'o' perms
 - 4 Unpriv user can only map own host uid
- 3 Other namespaces owned by a user ns
 - 1 Root in ns has full privilege over what it owns

Uid delegation

- 1 Root delegates *subuids* to users
 - 1 `/etc/subuid` and `/etc/subgid`: `serge:100000:65536`
 - 2 Set using `usermod`: `usermod -v 100000-200000 -w 100000-200000 serge`
- 2 Setuid-root programs write to `/proc/self/{ug}id_map`
- 3 Each user may be delegated a set of subuids and subgids

How safe is this?

- users including root in container are *unprivileged* users, but are local users.
- Linux Kernel CVEs
 - As of 2014-11-02, per <http://cvedetails.com>
 - 101 Total
 - 14 Priviledge Gain
- Use Cases
 - run existing service in container: +1

Distribution / Toolkit Support

- Linux kernel: added in 3.8, necessary improvements in 3.10
- Distros:
 - Ubuntu 14.04+
 - Red Hat Enterprise Linux: 7 (3.10.0-123 kernel)
 - SUSE Linux Enterprise Server: 12 (3.12.28-4.6)
- Tools:
 - lxc: in 1.0 improvements in 1.1
 - libvirt: yes, current versions
 - nova libvirt-lxc driver: Juno
 - nova-compute-flex: yes

More Info

- These Slides: <https://github.com/hallyn/ods-2014-usersns>
- Serge Hallyn <serge.hallyn@canonical.com> [freenode: 'hallyn']
- Scott Moser <scott.moser@canonical.com> [freenode: 'smoser']
- 'Namespaces In Operation' on lwn.net
- 10 part series on LXC 1.0
- LWN.net Secure Linux containers