Is it contained?

Untrusted root users in containers

Marco d'Itri
<md@linux.it>
@rfc1036

All Systems Go! 2018 - 29 September 2018

The experiment

The goal is to prevent an untrusted root user from escaping from a container to the host system.

The environment

- Debian stable (systemd 232, kernel 4.9.0)
- systemd-nspawn
- A normal (Debian) filesystem booted in a container

The configuration

```
# cat /etc/systemd/nspawn/test.nspawn
[Exec]
PrivateUsers=65536
DropCapability=CAP_AUDIT_CONTROL CAP_AUDIT_READ CAP_AUDIT_WRITE
 CAP_BLOCK_SUSPEND CAP_MKNOD CAP_NET_RAW CAP_SYS_ADMIN
 CAP_SYS_MODULE CAP_SYS_RAWIO CAP_SYS_TIME CAP_SYSLOG
 CAP_WAKE_ALARM

[Files]
TemporaryFileSystem=/run/lock
PrivateUsersChown=yes
```

What is left: CAP_CHOWN, CAP_DAC_OVERRIDE, CAP_DAC_READ_SEARCH, CAP_FOWNER, CAP_FSETID, CAP_IPC_LOCK, CAP_IPC_OWNER, CAP_KILL, CAP_LEASE, CAP_LINUX_IMMUTABLE, CAP_NET_ADMIN, CAP_NET_BIND_SERVICE, CAP_SETGID, CAP_SETFCAP, CAP_SETPCAP, CAP_SETUID, CAP_SYS_BOOT, CAP_SYS_CHROOT, CAP_SYS_NICE, CAP_SYS_PACCT, CAP_SYS_PTRACE, CAP_SYS_RESOURCE, CAP_SYS_TTY_CONFIG, CAP_SYSLOG.

Errors

```
systemd-nspawn[4770]: Failed to set hostname to <test-container>:
 Operation not permitted
systemd-nspawn[4770]: Failed to read AF_UNIX datagram queue
 length, ignoring: No such file or directory
systemd-nspawn[4770]: Failed to install release agent, ignoring:
 No such file or directory
systemd-nspawn[4770]: user.slice: Failed to set invocation ID on
 control group /user.slice, ignoring: Operation not permitted
...
```

The system appears to work.

But is it secure?

If not, could it be made secure with current kernel features?

Any questions?

`https://www.linux.it/~md/text/isitcontained-asg2018.pdf`

(Google . . . Marco d'Itri . . . I feel lucky)