Audit and Namespaces, future in Containers

Richard Guy Briggs
Senior Software Engineer, Red Hat
2016-08-25, Linux Security Summit

Who am I?

- Commodore PET 2001, Waterloo Structured BASIC, 1978
- PDP-11/23+ FORTRAN, 1987
- B.A.Sc. Comp.Eng, UOttawa, Linux 1992
- FreeS/WAN 1997
- Linux Imager drivers, 2007
- Red Hat kernel Audit since 2012
- “SunRaycer”, “RGB”, “papa”, “weird bike guy”

What is Audit?

- Intro: Rik Faith, Red Hat, 2004, 2.6.12-rc2
- Syslog on steroids
- Secure logging in the kernel
- Works well with SELinux
- Userspace daemon, log to disk or net
- Configurable kernel filters
- Only reports behaviour, not actively interfering

What are namespaces?

- Kernel-enforced userspace views
- peers:
  - MNT – custom filesystem view, 2001, 2.4.19
  - UTS – hostname, 2.6.19
  - IPC – InterProcess Communication, 2.6.19
  - NET – Independent network stacks, 2.6.24
- hierarchies
  - PID – Process IDs, 2.6.24 (each start at 1)
  - User – User IDs, 3.8 (unpriv. can clone, become 0 - root)
  - Cgroups – hides system limits (2016-03, 4.6)

What are containers?

- Many definitions
- Combo: namespaces, seccomp, cgroups
- Kernel has no concept
- Userspace container manager knows, reports
- ContainerID or collection of nsIDs

What's the problem?

- Highlander: “There can be only one”
- MNT, UTS, IPC namespaces no issues
- Wide open until 3.7-rc1
- NET – init NET ns only, fixed 3.14
- PID – init PID ns only, user msg fixed 3.15
- User – security concerns, ongoing
- Auditd – makes most sense tied to user ns

•NET – init NET ns only

- More than one proposal
- Least complex won out short term: 3.14
- Broke existing containers preventing login
- PAM broke assuming ECONNREFUSED when audit not available
- Non-init PID or User ns now returned EPERM
- We now lie in non-init user ns to unbreak

•PID – init PID ns only

- Audit user messages fixed 3.15
- Requires CAP_AUDIT_WRITE
- Vsftpd auth
- Cleaned up PID/PPID reporting
- Will allow CAP_AUDIT_CONTROL when user ns fixed

User ns

- Gao Feng: 2013
- Logical place for auditd
- AuditNS also proposed
- Still requires CAP_AUDIT_CONTROL
- Only one auditd per user ns
- Cannot influence init auditd config
- Own rulespace
- Own queue

NamespaceID

- 2013-03: Aristeu Rozanski, proc inode
- Ns serial # prototyped, discarded
- Reworked for nsfs with devID
- Each event includes set of nsIDs
- Audit logs aggregated by container orch.
- Container orch. keeps track across hosts

ContainerID

- Add to task struct (like sessionID)
- Each event includes ContainerID
- Set by Container orchestrator
- Inherited by children
- Audit logs aggregated by container orch.
- Container orch. keeps track across hosts

Conclusion

- Auditd ok with MNT, UTS, IPC ns
- NET ns ok for now
  - Will need audit_pid/portid per user ns
- PID ns ok for now for audit user messages
  - Will need translation per PID ns
- Auditd per user ns wanted for containers
- NamespaceID vs. ContainerID
- Need audit log aggregation by container orch.

Contact

- rgb@redhat.com
- Linux-audit@redhat.com
- Github.com/linux-audit