Linux Audit: Moving Beyond Kernel Namespaces to Audit Container IDs

Richard Guy Briggs
Senior Software Engineer, Red Hat
Linux Security Summit
2018-08-27&28, Vancouver

Who am I?

- CBM PET2001, 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”
- Humpty Dumpty

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/other LSMs
- Userspace daemon, log to disk or net
- Configurable kernel filters
- Only reports behaviour, not actively interfering

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”
- Audit can’t trace task to container
- Container security claims, tracking
- Required to filter logging and searches
- Route audit msgs to relevant daemon
- nsID tracking complex and incomplete

History

- 2013-03: Aristeu Rozanski, proc inode
- Added devID to qualify proc inode
- NS serial # prototyped, discarded
- Reworked for nsfs
- Each event includes set of nsIDs
- Abandon nsID as insufficient to ID containers
- NsID patchset updated v8, still potential use

LSS16: Conclusion

- Auditd ok with MNT, UTS, IPC, CGRP 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.

Changes since nsID proposal

- containers can't be universally identified by namespace (sub)set
- audit daemon won't be tied to any namespace
- netNS needs list of possible contIDs responsible for net events
- nsID info still potentially useful, but not pivotal
- Group audit task info into one struct (kABI)
- 3 revs of design, 4 revs of code

Access Controls

- can't unset contid
- must have CAP_AUDIT_CONTROL for w+r
- target task must not have threaded or spawned
- child inherits parent's contid
- possibly restrict to orchestrator's children
- disable setting twice?

Records/Fields

- u64 (u32, u128 and c36 considered)
- AUDIT_CONTAINER_OP record when created
- AUDIT_CONTAINER aux record to events, if contid set
- new field “contid”/AUDIT_CONTID for filtering
- add/del contid to netNS for net events, NETFILTER_PKT events

Remain to address

- how to allow multiple audit daemons
  - each will have its own queue and ruleset
  - auxiliaries can't influence host
- how to assign/route audit messages by contid
- LSM hooks to set the contid

Conclusion

- nsID infeasible to track containers
- u64 balances kernel efficiency and uniqueness
- Record for each of creation and event
- Filter by contid
- NET ns isolated events special treatment
- Audit logs aggregated by container orch.
- Container orch. keeps track across hosts

Contact

- rgb@redhat.com
- linux-audit@redhat.com
- github.com/linux-audit
- irc://FreeNode/#audit
- rgb@tricolour.ca
- irc://{OFTC,FreeNode}: @SunRaycer