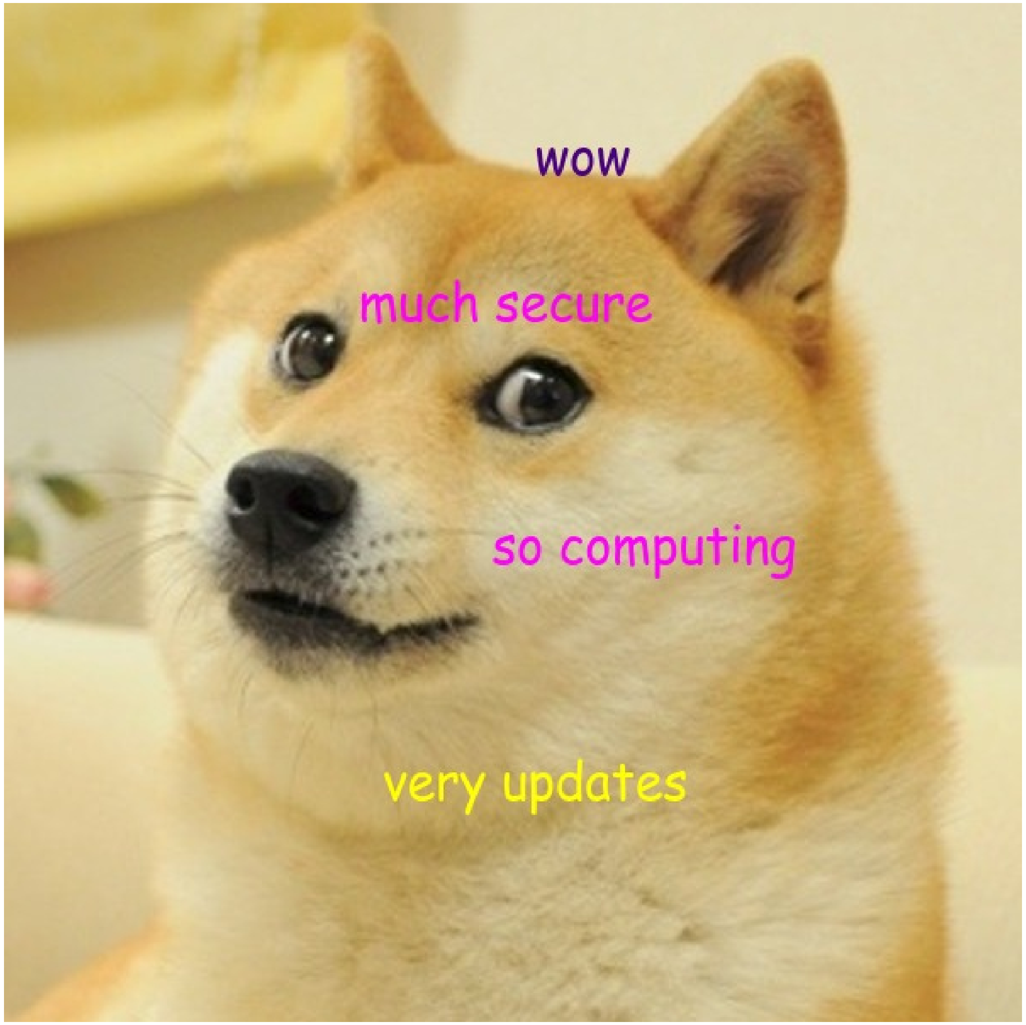
seccomp update



https://outflux.net/slides/2015/lss/seccomp.pdf

Linux Security Summit, Seattle 2015

Kees Cook <keescook@chromium.org>
(pronounced “Case”)

chrome

What is seccomp?

- Programmatic kernel attack surface reduction
- Used by:
  - Chrome
  - vsftpd
  - OpenSSH
  - Systemd (“SystemCallFilter=...”)
  - LXC (blacklisting)
  - … and you too! (easiest via libseccomp)

Architecture support

- x86: v3.5
- s390: v3.6
- arm: v3.8
- mips: v3.15
- arm64: v3.19, AKASHI Takahiro
- powerpc: linux-next (v4.3), Michael Ellerman

split-phase internals

- v3.19, Andy Lutomirski
- Splits per-architecture calls to seccomp into 2 phases
- Speeds up simple (no tracing) callers
- Only used on x86 so far

Regression tests

- v4.2: moved the 48 tests from github into the kernel: tools/testing/selftests/seccomp/
- Shows some interesting glitches with restart_syscall on arm (hidden) and arm64 (hidden, unless compat, then exposed)
- Gained big-endian support during powerpc port
- Added s390 seccomp support today

Minor changes

- v4.0: SECCOMP_RET_ERRNO capped at MAX_ERRNO
  - Avoid confusing userspace
- v4.1: asm-generic for seccomp.h
  - Easier architecture porting

Future

- Argument inspection
- CRIU (checkpoint/restore)
  - PTRACE_O_SUSPEND_SECCOMP with CAP_SYS_ADMIN: linux-next (v4.3), Tycho Andersen
  - Serialize dump/restore of filters.
- eBPF
  - Use maps or tail calls instead of balanced if/else trees for checking syscall numbers.

Questions?

https://outflux.net/slides/2015/lss/seccomp.pdf

@kees_cook

keescook@chromium.org

keescook@google.com

kees@outflux.net