Contribution ID: **439** Type: **not specified**

seccomp listeners for nested containers

Friday 12 December 2025 10:00 (30 minutes)

Currently, seccomp listeners (created via SECCOMP_FILTER_FLAG_NEW_LISTENER [1]) are limited to a single listener per process [2]. This becomes problematic in nested container scenarios – for example, when an outer LXC runtime intercepts the mknod syscall while an inner container runtime needs to hook sysinfo. Today, container runtimes often work around this by disabling seccomp listeners when they detect confinement (see [3]). I propose discussing possible approaches to support multiple or nested listeners, user-space API design, and their kernel-level implications.

I've submitted a patchset on LKML [4].

[1] https://github.com/seccomp/libseccomp/blob/9b9ea8e7a173b96e59fb21e8d461365110e7b4ef/src/system.c#L405C13-L405C45
[2] https://github.com/torvalds/linux/blob/fd94619c43360eb44d28bd3ef326a4f85c600a07/kernel/seccomp.c#L1926
[3] https://github.com/lxc/lxc/blob/faefb7b82878bec2398f52d8bbb78272d0f50dc5/src/lxc/seccomp.c#L1198
[4] https://lore.kernel.org/all/20251202115200.110646-1-aleksandr.mikhalitsyn@canonical.com/

Primary author: MIKHALITSYN, Aleksandr (Canonical)

Presenter: MIKHALITSYN, Aleksandr (Canonical)

Session Classification: Containers and checkpoint/restore MC

Track Classification: Containers and checkpoint/restore MC