



Contribution ID: 3

Type: **not specified**

## Sched-Ext: The BPF extensible scheduler class

**CFP closes on July 12th.**

### Overview

sched\_ext is a Linux kernel feature which enables implementing host-wide, safe kernel thread schedulers in BPF, and dynamically loading them at runtime. sched\_ext enables safe and rapid iterations of scheduler implementations, thus radically widening the scope of scheduling strategies that can be experimented with and deployed, even in massive and complex production environments.

sched\_ext was first sent to the upstream list as an RFC patch set back in November 2022. Since then, the project has evolved a great deal, both technically, as well as in the significant growth of the community of sched\_ext users and contributors.

### Discussions

Note that there is as-yet no particular order to the following discussions, with the exception that we'd ideally like for David to present first to open the session.

**David Vernet (void@manifold.com):** The current status and future potential of sched\_ext

- Opening the MC by giving the room an update on the latest status of sched\_ext, and getting discussions started with respect to future directions we could take, as well as how the community is growing.
- Depending on the outcomes of discussions at LSFMM, may discuss some current open questions as well; especially related to componentization. These may also be discussed by others, such as Andrea as described below.

**Andrea Righi (andrea.righi@canonical.com):** Crafting a Linux kernel scheduler that runs in user-space using Rust

- Necessary to coordinate how to properly componentize the scx repo to accommodate user space rust schedulers, rust hybrid schedulers, and C hybrid schedulers.
- The scheduler is interactive, so it will be an opportunity to discuss techniques that are appropriate for other schedulers such as scx\_lavd being experimented with to optimize the SteamDeck.

**Kevin Becker (kevin.becker@canonical.com):** Enabling sched-ext in the Ubuntu real-time kernel

- sched\_ext is thus far not thoroughly used or tested in PREEMPT\_RT kernels. We need to get distro maintainers in the same room as sched\_ext developers to discuss implications and plan for support.

**Giovanni Gherdovich (ggherdovich@suse.cz):** Enabling sched-ext in SUSE

- SUSE is interested in exploring the use of sched\_ext, and this would give SUSE distro maintainers an opportunity to discuss plans, use cases, and road blocks. This is especially useful given that other distro maintainers will be in the room.

**Piotr Górski (lucjan.lucjanov@gmail.com) and Peter Jung (admin@ptr1337.dev):** Deploying and managing sched\_ext schedulers in CachyOS

- CachyOS is a somewhat new, but powerful distribution that focuses on performance and scheduling. They were the first distribution to adopt sched\_ext, and want to share their plans for building a management layer and coordinating with other distros to avoid duplicated efforts.

**Andrea Righi (andrea.righi@canonical.com):** Distro-centric meeting session: solving generic toolchain and other dependency issues for deploying sched\_ext

- Leaving aside time to discuss general distribution problems with deploying sched\_ext, such as toolchain dependencies, backwards compatibility challenges, etc. There may be some overlap with other distro-centric discussions, but this discussion can be a way to ground them and develop action items that will apply to all distros.

**Changwoo Min (changwoo@igalia.com):** Using sched\_ext to improve frame rates on the SteamDeck

- Igalia has been working to leverage sched\_ext support on the SteamDeck, and has been seeing encouraging results. This discussion will allow us to discuss techniques that do and don't work for interactive workload testing, optimizing for gaming workloads, and how to validate interactive scheduler changes.

**Himadri Chhaya-Shailesh (himadrispandya@gmail.com):** Adaptive workload parallelization on over-subscribed hosts using sched\_ext

- A potential use case of sched\_ext is leveraging it for optimizing virtualization; both with paravirt, or with no guest awareness. Himadri has been experimenting with this, and can discuss pain points she's encountered, and roadmap items that will be needed to further enable this effort.

#### Logistics notes

- This is the first time we've proposed a sched\_ext microconference, so we have no results and accomplishments to discuss from prior meetings.
- We've discussed the possibility of combining with either the sched or sched-RT microconferences with the authors of those MCs, and we all agree that there are already too many topics to combine the MCs. That said, if at all possible, it would be great if we could schedule the scheduler-related MCs such that they don't overlap, so that we can attend sessions in the other MCs when possible.
- We would ideally like to allow around 30 minutes per presentation, if at all possible, so around 4 hours + break times. If this is too long, let us know and we can accommodate.

**Primary author:** VERNET, David (Meta)

**Co-authors:** RIGHI, Andrea (Canonical); MIN, Changwoo (Igalia); Mr GHERDOVICH, Giovanni (SUSE); Mrs CHHAYA-SHAILESH, Himadri (Inria-Paris); Mr BECKER, Kevin (Canonical); Mr JUNG, Peter (CachyOS); Mr GÓRSKI, Piotr (CachyOS)

**Presenters:** RIGHI, Andrea (Canonical); MIN, Changwoo (Igalia); VERNET, David (Meta); Mr GHERDOVICH, Giovanni (SUSE); Mrs CHHAYA-SHAILESH, Himadri (Inria-Paris); Mr BECKER, Kevin (Canonical); Mr JUNG, Peter (CachyOS); Mr GÓRSKI, Piotr (CachyOS)

**Track Classification:** LPC Microconference Proposals