



Contribution ID: 87

Type: **not specified**

cgroup rstat's advanced adoption

Wednesday, 14 September 2022 11:45 (25 minutes)

rstat is a framework how generic hierarchical stats collection is implemented for cgroups.

It is light on the writer (update) side since it works with per-cgroup per-cpu structures only (mostly).

It is quick on the reader side since it aggregates only cgroups active since the previous read in a given subtree.

It is used for accounting CPU time on the unified hierarchy, blkcg and memcg stats. Readers of the first two are user space queriers, the memcg stats are used additionally by MM code internally and hence memcg builds some optimizations above rstat. Despite that there were reports of readers being negatively affected by occasionally too long stats retrieval.

This is suspected to be caused by some shared structures within rstat and their effect may get worse as more subsystems (or even BPF) start building upon rstat.

This talk describes how rstat currently works and then analyzes time complexity of updates and readings depending on number of active use sites.

The result could already be a base for discussion and we will further consider some approaches to keep rstat durations under control with more new adopters and also how such methods affect error of collected stats (when tolerance is limited, e.g. for the VM reclaim code).

This presentation and discussion will fit in a slot of 30 minutes (give or take).

I agree to abide by the anti-harassment policy

Yes

Primary author: KOUTNÝ, Michal

Presenter: KOUTNÝ, Michal

Session Classification: Containers and Checkpoint/Restore MC

Track Classification: LPC Microconference: Containers and Checkpoint/Restore MC