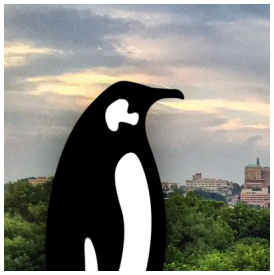

Contribution ID: **39** Type: **not specified**

Improve Linux Perf tool to account for task sleep

Tuesday, 14 November 2023 10:15 (45 minutes)

Problem: As per the current architecture of Linux Perf tool, 'perf record'does not collect samples if target process is in sleep state. Due to this perf tool has following limitations:

Incorrect 'CPU usage'calculation: If target task was in sleep state for around 50% of the time, the CPU usage represented by perf tool does not account for the same.

No 'task sleep time': As perf tool does not provide any sleep sample, so it's not possible to determine for how long the task was in sleep state.

Solutions: Perf-record sampling happens when perf_swevent_hrtimer() handler executes. If the target process is in sleep state, the handler is not being called.

1) When perf_swevent_hrtimer() handler executes, it can calculate missing samples for the period when the target was in sleep state, using:
missed_sample_count = ((current_time –hrtimer_start_time) / sampling_freq)
missed sample count would have to be sent to user space perf-sample handler which stores this information to perf.data. And perf-report processes all missed samples and adds them to total samples.

2) User space perf tool could calculate CPU usage based upon expected samples instead of total samples collected, as shown:
expected_sample = total_time / freq

3) Change the behaviour of perf_swevent_hrtimer() handler so that it should always be called even if target task is in sleep state (either wake up the target task or run in another task's context).

Primary authors: KAHER, Ajay; MAKHALOV, Alexey

Presenters: KAHER, Ajay; MAKHALOV, Alexey

Session Classification: Birds of a Feather (BoF)

Track Classification: Birds of a Feather (BoF)