

PSI monitor for memory
pressure detection

Problem

Main issues with vmpressure signals for memory pressure detection:

- Reflects current reclaim efficiency rather than memory pressure level
- Difficult to tune because of no direct link between reclaim efficiency and its effects on user experience
- Tightly coupled with vmscan implementation, changes in vmscan mechanisms may result in behavior change
- In testing highly depends on the system memory size and particular workload

As a result even low vmpressure signals have to be monitored and additional hints should be used to filter out false positives.

PSI (Pressure Stall Information)

PSI mechanism introduced in v4.20-rc1 release measures amount of time tasks are delayed as a result of resource contention. It is recorded as total stall time and averages over 10, 60 and 300 secs are provided.

Contended resources are: Memory, IO and CPU.

PSI separately records complete (FULL) and partial (SOME) stall amounts.

Considers number of CPUs and number of non-idle tasks.

Represents a more direct measure of how resource contention affects user experience.

More details at: <https://facebookmicrosites.github.io/psi/docs/overview>

Why PSI averages are not enough for Android

Mobile systems have:

- relatively limited amount of memory
- applications and services with unpredictable and often aggressive memory requirements

Averages over multiple seconds are useful for monitoring long-term trends but will grow too slow to detect and react to spikes in memory usage. System becomes unusable before appropriate measures can be deployed.

Userspace could poll total stall metrics but doing that continuously at a relatively high frequency would have high performance cost.

PSI monitor

Heavily utilizes already existing PSI machinery to minimize overhead

API similar to vmpressure allowing userspace to wait for events (select/poll/epoll)

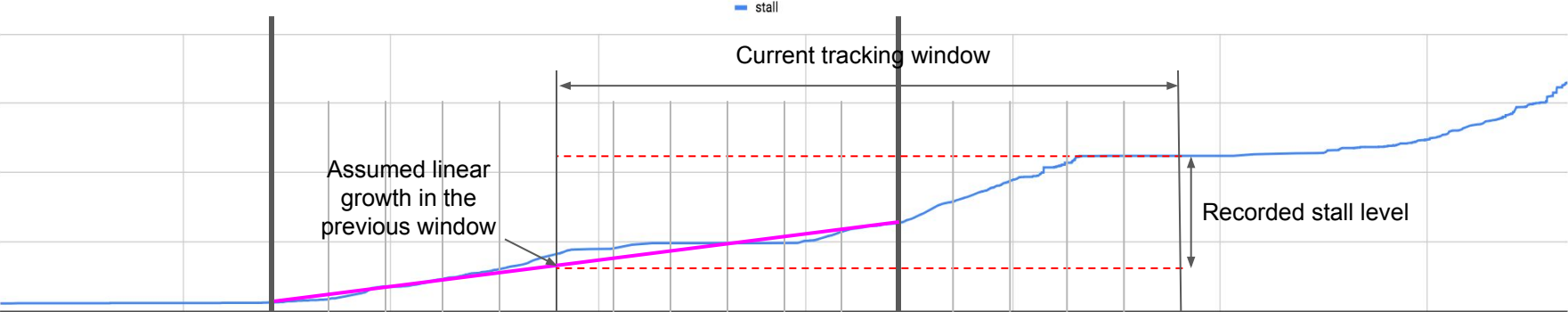
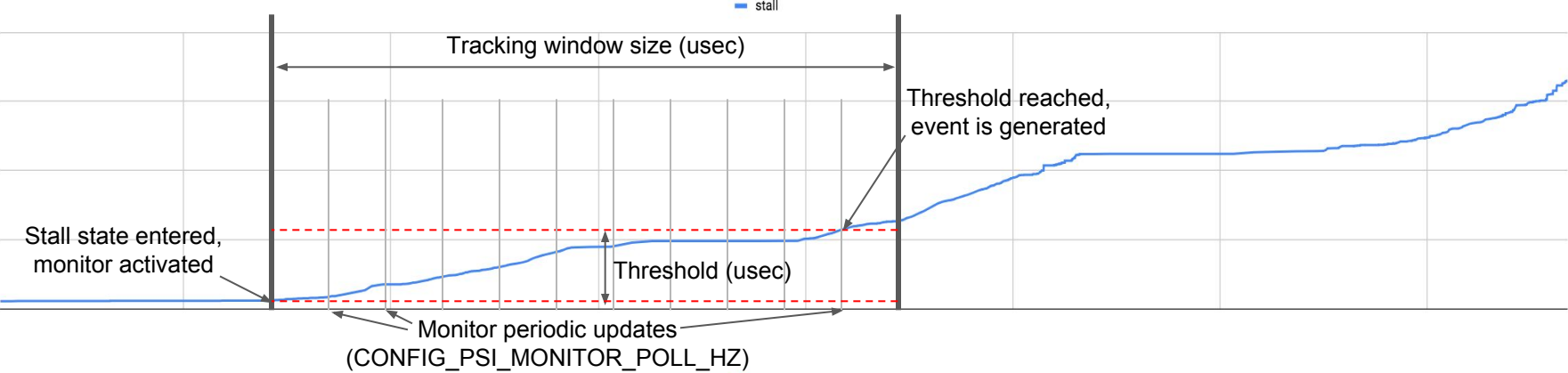
Activates only when monitored metric enters stall state and deactivates upon exit

Polls stall duration at a frequency configurable via kernel configuration flag

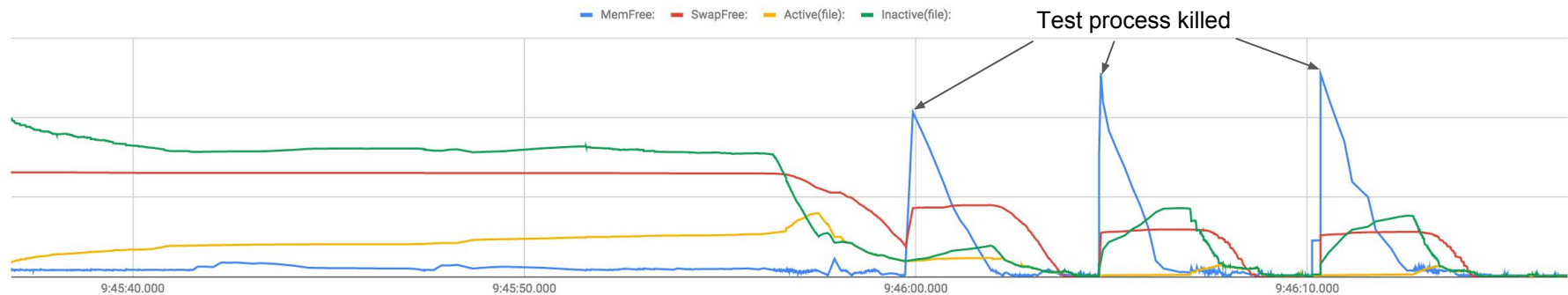
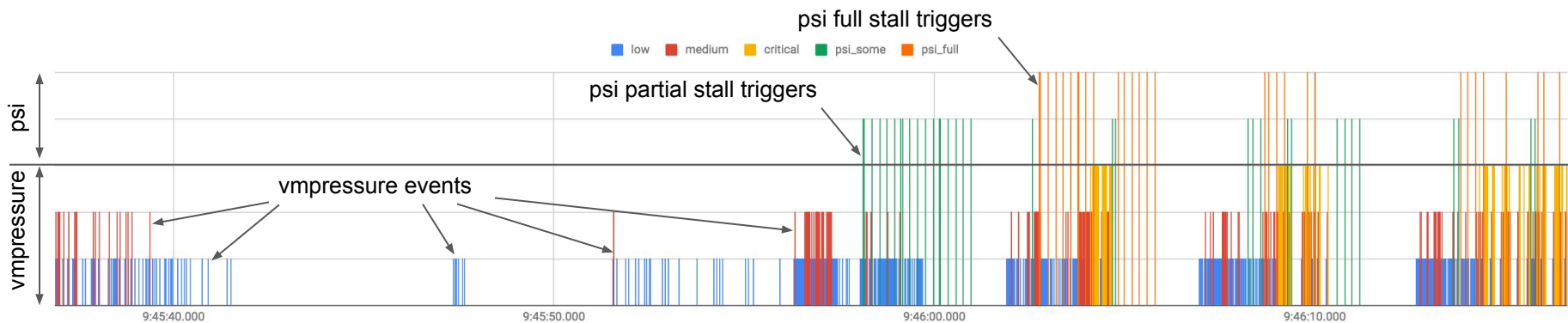
Supports multiple concurrent monitors for all PSI metrics (SOME / FULL combined with Memory / IO / CPU)

Provides per-monitor configurable threshold and tracking window size

PSI monitor implementation



psi monitor vs vmpressure during memory stress test



PSI monitor configured to trigger at 10% partial stall and 7% full stall using 1sec tracking window

Discussion