Contribution ID: **100** Type: **not specified**

Finding more DRAM

Wednesday 11 September 2019 10:00 (45 minutes)

The demand of DRAM across different platforms is increasing but the cost is not decreasing. Thus DRAM is a major factor of the total cost across all kinds of devices like mobile, desktop or servers. In this talk we will be presenting the work we are doing at Google, applicable to Android, Chrome OS and data center servers, on extracting more memory out of running applications without impacting performance.

The key is to proactively reclaim idle memory from the running applications. For the Android and Chrome OS, the user space controller can provide hints of the idle memory at the applications level while the servers running multiple workloads, an idle memory tracking mechanism is needed. With such hints the kernel can proactively reclaim memory given that estimated refault cost is not high. Using in-memory compression or second tier memory, the refault cost can be reduced drastically.

We have developed and deployed the proactive reclaim and idle memory tracking across Google data centers [1]. Defining idle memory as memory not accessed in the last 2 mins, we found 32% idle memory across data centers and we were able to reclaim 30% of this idle memory, while not impacting the performance. This results in 3x cheaper memory for our data centers. 98% of the applications spend only around 0.1% of their CPU on memory compression and decompression. Also the idle memory tracking on average takes less than 11% of a single logical CPU.

The cost of proactive reclaim and idle memory tracking is reasonable for the data centers cost of ownership of memory, however, it imposes challenges for power constrained devices based on Android and Chrome OS. These devices run diverse applications e.g. Chrome OS can run Android and Linux in a VM. To that end, we are working on making idle memory tracking and proactive reclaim feasible for such devices. Henceforth, we are interested and would like to initiate discussion on making proactive reclaim useful for other use-cases as well.

[1] Software-Defined Far Memory in Warehouse-Scale Computers, ACM ASPLOS 2019.

I agree to abide by the anti-harassment policy

Yes

Primary authors: BUTT, Shakeel (Google); BAGHDASARYAN, Suren (Google); ZHAO, Yu (Google)

Presenters: BUTT, Shakeel (Google); BAGHDASARYAN, Suren (Google); ZHAO, Yu (Google)

Session Classification: LPC Refereed Track