Contribution ID: **321** Type: **not specified**

Accelerating Linux Kernel Boot-Up for Large Multi-Core Systems

Wednesday 18 September 2024 10:40 (20 minutes)

The Linux kernel has been observed to take several 10s of seconds to boot-up on machines with many CPUs (~1792 CPUs). This talk delves into the details of bottlenecks uncovered in the CPU online path when testing on large NUMA multi-core virtual machines and outlines some of the fixes that helped achieve up to 50% faster boot times on such VMs. These optimizations range from approaches such as amortizing the cost of certain repetitive calculations by deferring them until all CPUs are up [1], to rewriting CPU hotplug callbacks as worker functions and leveraging the kworker infrastructure to run these callbacks in parallel on all the online CPUs.

Further, this talk will draw focus on the internals of the CPU hotplug framework, whose callback invocation is still primarily sequential and executes them one after another on a single CPU, irrespective of how many CPUs are already online (and thus readily available for parallel execution). This design gets particularly expensive for those CPU online callbacks whose computation involves loops (or nested loops with NUMA nodes) that span every online CPU. As a result, the current design incurs a linear degradation (or worse) in the execution time for such callbacks as the number of CPUs (and NUMA nodes) grows, thus making each CPU online operation progressively slow as the Linux kernel makes its way through the boot-up sequence.

We will discuss approaches to address these issues to scale booting and CPU online operations for large multi-core systems and seek to brainstorm with the community and get their invaluable feedback.

References:

[1]. [PATCH] mm/vmstat: Defer the refresh_zone_stat_thresholds after all CPUs bringup - Saurabh Sengar (kernel.org)
https://lore.kernel.org/all/1720169301-21002-1-git-send-email-ssengar@linux.microsoft.com/

Primary authors: SINGH SENGAR, Saurabh (Microsoft); BHAT, Srivatsa (Microsoft)

Presenters: SINGH SENGAR, Saurabh (Microsoft); BHAT, Srivatsa (Microsoft)

Session Classification: System Boot and Security MC

Track Classification: System Boot and Security MC