Contribution ID: **44**   Type: **not specified**

1G page support for guest_memfd

Thursday, 19 September 2024 12:00 (30 minutes)

guest_memfd is a new feature providing a guest-first memory subsystem internal to KVM. Being internal to KVM opens guest_memfd up to virtualization-specific features, enhancements and optimizations.

A notable feature that guest_memfd currently lacks is 1G page support.

Here are some key benefits of 1G page support:
Better performance due to
Increased hit rate in the TLB
Faster page table walks in case of TLB misses
Memory savings due to smaller page table
Further memory savings due to Vmemmap optimization

Support for 1G pages is crucial to support large virtual machines, which are becoming more common due to growing usage of GPU and other AI accelerator hardware such as TPUs.

Specifically for Confidential Virtual Machines (CVMs), there are further challenges in using 1G pages as guests tend to require conversion of memory ranges between two memory classes, private and shared, at granularities smaller than 1G.

We previously explored providing 1G pages from the hugetlb subsystem because the hugetlb subsystem already provides useful features such as a pool of huge pages to the kernel at boot time, together with page accounting, memory charging and reporting.

In this presentation, we want to
Present some known requirements (and hope to gather more) of 1G page support in guest_memfd, potentially including
Sharing of 1G page pool with other subsystems
Setting up pages for host CPU/IO access to guest_memfd
Explore and compare options for adding 1G page support including using the hugetlb subsystem
Explain challenges of using guest_memfd in CVMs and possible solutions
Discuss the best way of landing this support in the kernel

Primary authors: Mr TNG, Ackerley; ANNAPURVE, Vishal

Presenter: ANNAPURVE, Vishal

Session Classification: KVM Microconference

Track Classification: KVM Microconference