Contribution ID: **195** Type: **not specified**

Cooperative DMA in a memory-oversubscribed environment

Monday, 12 September 2022 10:30 (30 minutes)

Running virtual machines with memory subscription and DMA device passthrough is a challenge:
1. If devices/IOMMUs don't support faults or ATS, the hypervisor can't know which pages to map to ensure that DMA will not fault.
2. VFIO pins all memory when the memory range is mapped for DMA; this makes overcommit a challenge!

We describe a solution to both of these problems:
- support VFIO DMA (re)mapping: when a page is reclaimed via madvise or swap, remove it from IOMMU page table mappings; when a page is faulted in add it to IOMMU mappings. Similar to how KVM page tables are kept in sync with userspace page tables
- provide an light-weight enlightenment to virtual machine kernels which can cooperate with the hypervisor to ensure that pages mapped for DMA are resident

The overview of this solution is presented, and some open questions are posed for consideration by the audience:
- how to connect IOMMU page tables to userspace page tables? Callbacks?
- how to expose the DMA cooperative device to the guest virtual machine (or process)

Finally discussion about next steps and a path to upstreaming is discussed.

I agree to abide by the anti-harassment policy

Yes

Primary author: GOWANS, James (Amazon EC2)

Presenter: GOWANS, James (Amazon EC2)

Session Classification: VFIO/IOMMU/PCI MC

Track Classification: LPC Microconference: VFIO/IOMMU/PCI MC