IOMMU Page Faulting and MM Integration

Joerg Roedel <jroedel@suse.de>

Why do we need MM Integration?

- New PCI Hardware with PRI and PASID support
  - Fault recovering
  - Multiple address spaces per device
- Allows devices to directly access process address spaces
- Translation happen in the IOMMU
- IOMMU driver needs to setup mappings for the devices

Current State

- Hardware available from AMD
  - IOMMUv2
  - Newer Radeon GPU in APUs
- Support implemented in the AMD IOMMUv2 driver
  - Extension module to the built-in AMD IOMMU driver
  - Implements the page-fault loop for devices
- Currently pending an MMU notifier extension to fix an outstanding issue
  - Will send new version when 3.18-rc1 is out

Required MMU_Notifier Change

- MMU_Notifiers not suitable for page-table sharing
  - We need the remote-TLB flush event
  - All mmu_notifiers provide is invalidate_range_start/end
  - Wrong semantics
- Patch set under review to add an invalidate_range notifier
  - Will close this gap
  - Notifies about the remote-TLB flush event
  - Also notifies when page-table pages are freed

Future

- More Hardware with these capabilities is coming up
  - Intel SVM already specified in the Vt-d specification
  - Other architectures, non-pci?
- The existing AMD code needs to be turned into a generic IOMMU-API extension
- Code is mostly generic already, the call-backs into the in-kernel AMD IOMMU driver needs to be generallized
- Revisit PASID allocaiton/handling (David?)

Current Exported API

- int **amd_iommu_init_device**(struct pci_dev *pdev, int pasids)
- void **amd_iommu_free_device**(struct pci_dev *pdev)
- int **amd_iommu_bind_pasid**(struct pci_dev *pdev, int pasid,
  struct task_struct *task)
- void **amd_iommu_unbind_pasid**(struct pci_dev *pdev, int pasid)
- int **amd_iommu_set_invalid_ppr_cb**(struct pci_dev *pdev,
  amd_iommu_invalid_ppr_cb cb)
- int **amd_iommu_set_invalidate_ctx_cb**(struct pci_dev *pdev,
  amd_iommu_invalidate_ctx cb)

Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org